(19) (11) Publication number: **0**

Generated Document.

PATENT ABSTRACTS OF JAPAN

(21) Application number: **05051663**

(22) Application date: **12.03.93**

(51) Intl. Cl.: **G06F 3/06** G06F 3/06 G06F

(30) Priority:

(43) Date of application publication: **22.09.94**

(84) Designated contracting states:

(71) Applicant: **HITACHI LTD**

(72) Inventor: **TSUNODA HITOSHI**
TAKAMOTO YOSHIFU
KAMO YOSHIHISA

(74) Representative:

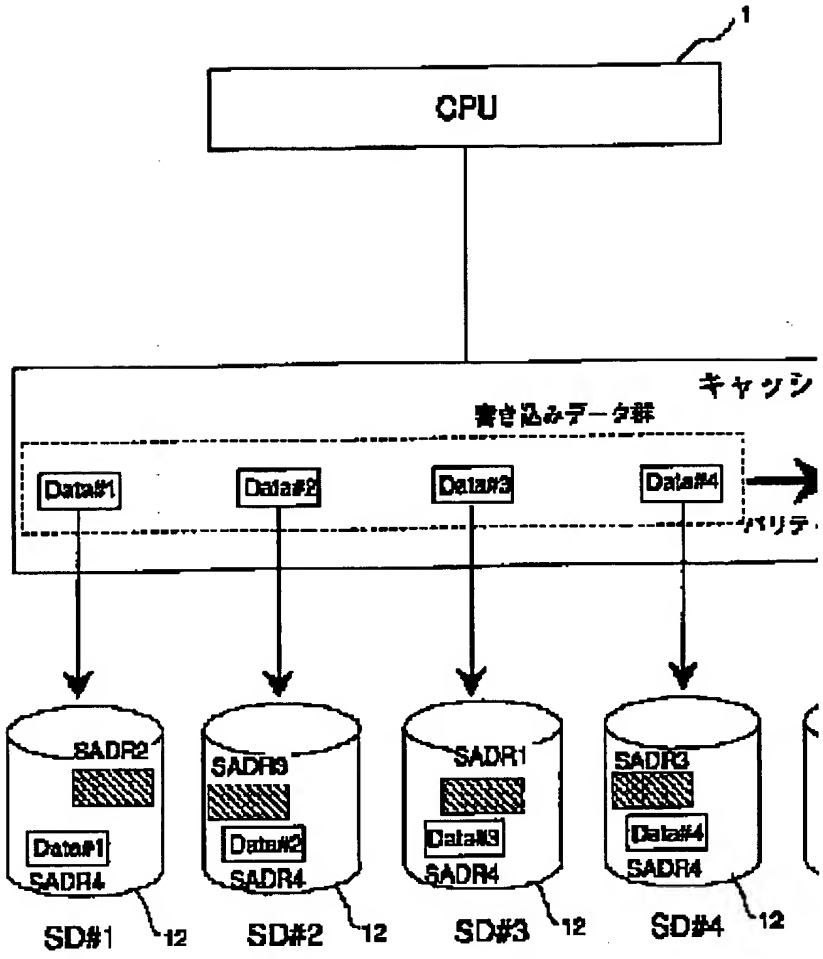
(54) DISK ARRAY SYSTEM AND DATA WRITE METHOD AND FAULT RECOVERY METHOD FOR THIS SYSTEM

(57) Abstract:

PURPOSE: To reduce the overhead for write with respect to a disk array of RAID (level 5) where data is distributed to improve the processing performance.

CONSTITUTION: Even if data #1 to #4 already written in addresses SADR1 to SADR3 in a drive as data belonging to groups different from one another will be rewritten with write data, these write data are regarded as new write data and are written in the idle area of an address SADR4 in the drive in parallel. Updated old data is not read out. A nullity flag is registered in an address conversion table with respect to updated old data, and data is read from the newly written area. When all of data in original parity groups are

made ineffective, areas holding these groups are used as idle areas. Effective data in parity groups which are made partially ineffective are justified at a proper timing.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平6－266510

(43)公開日　平成６年(1994)９月22日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	３０５　Ｃ	7165－5B		
	３０１　Ｚ	7165－5B		
	３０２　Ａ	7165－5B		

審査請求　未請求　請求項の数28　ＯＬ　（全 26 頁）

(21)出願番号　特願平5－51663

(22)出願日　平成５年(1993)３月12日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　角田　仁
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　高本　良史
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　加茂　善久
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内

(74)代理人　弁理士　小川　勝男

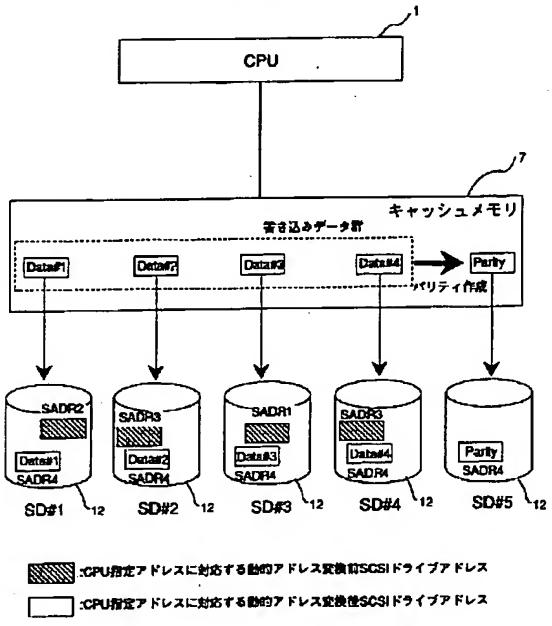
(54)【発明の名称】　ディスクアレイシステムおよびそのためのデータ書き込み方法、障害回復方法

(57)【要約】　（修正有）

【目的】　データを分散させて処理性能を向上させるＲＡＩＤ（ＬＥＶＥＬ５）のディスクアレイにおいて、書込み時のオーバヘッドを減少させる。

【構成】　書込みデータが、すでにドライブ内アドレスＳＡＤＲ２～ＳＡＤＲ３にそれぞれ異るグループに属するデータとして書込み済みのData#1#4を書換えるものであっても、新書込みデータと見なして、ドライブ内のアドレスSADR4の空領域に並列書込む。更新された旧データは読出さない。更新された旧データに関して無効フラグをアドレス変換テーブルに登録し新に書込まれた領域から行なう。元のパリティグループ内のすべてのデータが無効にされた場合、そのグループを保持していた領域は、空領域として使用する。部分的に無効にされたパリティグループ内の有効データは、適当なタイミングで詰め替える。

図6



1

【特許請求の範囲】

【請求項1】複数のディスクドライブと、一つまたは複数の上位装置から転送された、それぞれ該複数のディスクドライブに新に書き込まれるべき所定の長さのデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定長を有するデータのいずれかからなる複数の書き込みデータを一時的に保持するキャッシュメモリとを有するディスクアレイシステムにおいて、

(a) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成し、

(b) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込み、

(c) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にするステップからなるデータ書き込み方法。

【請求項2】(d) 該一つまたは複数の上位装置から転送された書き込みデータの量が、該レコード長を有する該所定数に等しい数の書き込みデータを取り出すのに十分になる毎に、上記ステップ(a)から(c)を実行するステップからなるデータ書き込み方法。

【請求項3】該誤り訂正データの生成は、

該一つの上位装置から、一つの書き込みデータが転送されたとき、その一つの書き込みデータをキャッシュ記憶に保持し、

該一つの上位装置から、該一つの書き込みデータが転送されたときに、該キャッシュに既に保持された他の書き込みデータの量が、該一定長を有する該所定数に等しい数の書き込みデータを取り出すのに十分であるときには、該一つまたは複数の上位装置から他の書き込みデータがそのキャッシュに転送されるのを待たないで、上記該誤り訂正データの生成を実行し、

該一つの上位装置から、該一つの書き込みデータが転送されたときに、該キャッシュに保持された書き込みデータの量が、該一定長を有する該所定数に等しい数の書き込みデータを取り出すのに十分でないときには、該キャッシュに保持された書き込みデータの数が、該所定数に達した後、上記該誤り訂正データの生成を実行するステップからなる請求項1記載のデータ書き込み方法。

【請求項4】(d) 該ステップ(a)から(c)を、該キャッシュメモリに保持された、他の書き込みデータに関して繰り返し、

(e) 該ステップ(d)を繰り返し実行した結果、該複数の

2

ドライブに記憶されていたいずれか一つの誤り訂正データグループに属する一群の書き込みデータがすべて無効となったとき、その後のステップ(d)によるステップ(a)から(c)の繰り返し特に、その一群の書き込みデータおよびその誤り訂正データグループに属する誤り訂正用データを保持していた、該複数のドライブ内の複数の領域を、空き領域としていずれか他の誤り訂正データグループの書き込みに使用するステップをさらに有する請求項一記載のデータ書き込み方法。

【請求項5】複数のディスクドライブと、一つまたは複数の上位装置から転送された、それぞれ該複数のディスクドライブに新に書き込まれるべき所定の長さを有するデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定長を有するデータのいずれかからなる複数の書き込みデータを一時的に保持するキャッシュメモリとを有するディスクアレイシステムにおいて、

(a) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正を用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成し、

(b) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込み、

(c) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にするステップからなるデータ書き込み方法。

(d) 該ステップ(a)から(c)を、該キャッシュメモリに保持された、他の書き込みデータに関して繰り返し、

(e) 該複数のドライブの一つに障害が発生したとき、その一つのドライブに保持されていた複数のデータのうち、複数の有効なデータの各々を、該一つのドライブ以外のドライブに保持されている、そのデータが属する誤り訂正データグループに属する、複数の他の有効なあるいは無効な書き込みデータと誤り訂正用データとを用いて選択的に回復するステップからなる障害回復方法。

【請求項6】該障害の発生前に、いずれかの一群の書き込みデータに対して、該ステップ(c) を実行した結果、そのステップ(c) で無効なデータとされたいずれか一つのデータが属する誤り訂正データグループに属する一群の書き込みデータがすべて無効なデータとなった場合には、その一群の書き込みデータおよびその誤り訂正グループに属する誤り訂正用データを保持していた、該複数のドライブ内の複数の領域を、空き領域として他の誤り訂正データグループに属するデータの書き込みに使用するステップをさらに有する請求項5記載の障害回復方

3

法。

【請求項7】該回復ステップは、

該複数のドライブの内、該一つのドライブ以外のドライブより、該複数の他の有効なまたは無効な書き込みデータと該誤り訂正用データとを、選択的に読み出し、

該読み出された書き込み済みデータと該読み出された誤り訂正用データとから、該一つの有効なデータを回復するステップからなる請求項5記載の障害回復方法。

【請求項8】複数のディスクドライブと、一つまたは複数の上位装置から転送された、それぞれ該複数のディスクドライブに新に書き込まれるべき所定の長さのデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定長を有するデータのいずれかからなる複数の書き込みデータを一時的に保持するキャッシュメモリとを有するディスクアレイシステムにおいて、

(a) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正を用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成し、

(b) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込み、

(c) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にするステップからなるデータ書き込み方法。

(d) 該ステップ(a)から(c)を、該キャッシュメモリに保持された、他の書き込みデータに関して繰り返し、

(e) それぞれ少なくとも一つの有効な書き込みデータと少なくとも一つの無効な書き込みデータを含む複数の誤り訂正データグループ（部分的に無効な誤り訂正データグループ）に含まれる複数の有効な書き込みデータを詰め替えるステップからなり、その詰め替えは、(e1) 部分的に無効な誤り訂正データグループに含まれている一つまたは複数の有効な書き込みデータを、該複数のドライブから選択的に読み出し、読み出された一つまたは複数の有効な書き込みデータをキャッシュ記憶の保持し、(e2) 該ステップ(e1)を他の部分的に無効な誤り訂正データグループに対して実行し、(e3) 該キャッシュ記憶に保持された、複数の有効な書き込みデータから選択された該所定数に等しい数の有効な書き込みデータの新たな組みとその書き込みデータの組みに対して新に誤り訂正用データを生成し、もって、新たな誤り訂正データグループを生成し、(e4) 該生成された新たな誤り訂正データグループを該複数のドライブ内の空き領域に並列に再度書き込み、(e5) 該ステップ(e3)(e4)を他の新たな誤り訂正データグループのために繰返し、(e6) 複数の部分無効誤り訂正データグループであって、そこに含まれていた有効な書き込みデータが該ステップ(e4)(e5)により、すべて再書き込みされたものを保持していた複数の領域を、空き領域として他の誤り訂正データグループの書き込みに使用するステップを有するデータ書き込み方法。

4

【請求項9】上記ステップ(e1)または(e2)は、特定状態にある複数の部分的に無効な誤り訂正データグループに対して選択的に行なわれる請求項8記載のデータ書き込み方法。

【請求項10】該特定の状態は、部分的に無効な誤り訂正データグループに含まれた無効とされた書き込みデータの数が、予め定めた限界数以上である状態である請求項8記載のデータ書き込み方法。

【請求項11】上記ステップ(e1)または(e2)は、

いずれか一つの誤り訂正データグループが該特定状態に達するごとに、その誤り訂正データグループに対して行なわれる請求項9記載のデータ書き込み方法。

【請求項12】上記ステップ(e1)または(e2)は、

いずれか一つの該特定状態にある誤り訂正データグループに属する有効な書き込みデータ対していずれか一つの上位装置から転送された読み出し要求に応答してその要求された書き込みデータに対してなされる請求項9記載のデータ書き込み方法。

【請求項13】上記ステップ(e1)または(e2)は、

複数の部分的に無効な誤り訂正データグループの数が、所定数に達したときに、該複数の部分的に無効な誤り訂正データグループに対して行なわれる請求項10記載のデータ書き込み方法。

【請求項14】上記ステップ(e1)または(e2)は、

該複数のドライブに含まれたデータ書き込み用の領域の内の空き領域の容量が、予め定めた限界値に達したときに、複数の部分的に無効な誤り訂正データグループに対して選択的に行なわれる請求項9記載のデータ書き込み方法。

【請求項15】上記ステップ(e1)または(e2)は、

該複数のドライブに含まれたデータ書き込み用の領域の容量に対する空き領域の容量の比率が、予め定めた限界比率に達したときに、複数の部分的に無効な誤り訂正データグループに対して選択的に行なわれる請求項9記載のデータ書き込み方法。

【請求項16】上記一つまたは複数の上位装置から新たな書き込みデータが転送されたとき、該書き込みデータを該キャッシュ記憶に書き込むステップをさらに有し、

上記ステップ(e3)は、該キャッシュにいずれかの上位装置から転送された新たな書き込みデータがあるときには、その新たな書き込みデータと該ステップ(e1)で読み出された書き込みデータとからなる該所定数に等しい数の書き込みデータを含む誤り訂正データグループを生成

5

するステップをさらに有する請求項9記載のデータ書き込み方法。

【請求項17】複数のディスクドライブと、一つまたは複数の上位装置から転送された、それぞれ該複数のディスクドライブに新に書き込まれるべき所定の長さのデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定の長さを有するデータのいずれかからなる複数の書き込みデータを一時的に保持するキャッシュメモリとを有するディスクアレイシステムにおいて、

(a) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正を用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成し、

(b) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込み、

(c) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にするステップからなるデータ書き込み方法。

(d) 該ステップ(a)から(c)を、該キャッシュメモリに保持された、他の書き込みデータに関して繰り返し、

(e) それぞれ少なくとも一つの有効な書き込みデータと少なくとも一つの無効な書き込みデータを含む誤り訂正データグループ（部分的に無効な誤り訂正データグループ）に含まれる複数の有効な書き込みデータを詰め替えるステップからなり、その詰め替えは、(e1) いずれか一つの上位装置から転送された、いずれか一つの部分的に無効な誤り訂正データグループに含まれている有効な書き込みデータに対する読み出し要求に応答して、その書き込みデータを該複数のドライブから選択的に読み出し、かつ、該読み出された書き込みデータを、該一つの上位装置に転送するとともに、キャッシュ記憶に保持し、(e2) 該ステップ(e1)および(e2)を他の部分的に無効な誤り訂正データグループに対して実行し、(e3) 該キャッシュ記憶に保持された、複数の書き込みデータから選択された該所定数に等しい数の書き込みデータの新たな組みとその書き込みデータの組みに対して新に誤り訂正用データを生成し、もって、新たな誤り訂正データグループを生成し、(e4) 該生成された新たな誤り訂正データグループを該複数のドライブ内の空き領域に並列に書き込み、(e5) 該ステップ(e1)または(e2)により読み出された書き込みデータが、該ステップ(e4)により再書き込みされたとき、該複数のドライブに保持されている、その再書き込みされる前のその書き込みデータを無効にするステップを有するデータ書き込み方法。

6

【請求項18】上記ステップ(e3)から(e4)は、該ステップ(e1)または(e2)で読み出された書き込みデータが属する部分的に無効な誤り訂正データグループが特定状態にあるときに行なわれる請求項17記載のデータ書き込み方法。

【請求項19】該特定の状態は、部分的に無効な誤り訂正データグループに含まれた無効とされた書き込みデータの数が、予め定めた限界数以上である状態である請求項18記載のデータ書き込み方法。

【請求項20】複数のディスクドライブと、一つまたは複数の上位装置から転送された、それぞれ該複数のディスクドライブに新に書き込まれるべき所定長を有するデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定長のデータのいずれかからなる複数の書き込みデータを一時的に保持するキャッシュメモリとを有するディスクアレイシステムにおいて、

(a) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正を用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成し、

(b) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込み、

(c) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にするステップからなるデータ書き込み方法。

(d) 該ステップ(a)から(c)を、該キャッシュメモリに保持された、他の書き込みデータに関して繰り返し、

(e) それぞれ少なくとも一つの有効な書き込みデータと少なくとも一つの無効な書き込みデータを含む誤り訂正データグループ（部分的に無効な誤り訂正データグループ）に含まれる複数の有効な書き込みデータを詰め替えるステップからなり、その詰め替えは、(e1) いずれか一つの部分的に無効な誤り訂正データグループに含まれている一つまたは複数の有効な書き込みデータの少なくとも一つを該複数のドライブから選択的に読み出し、かつ、該読み出された書き込みデータを該キャッシュ記憶に一時的に保持し、(e2)いずれか一つの上位装置から新たな書き込みデータが転送されてきたときに、その書き込みデータを一時的に該キャッシュ記憶に保持し、(e3) 該ステップ(e1)または(e2)により該キャッシュ記憶に保持された、複数の書き込みデータから該所定数に等しい数の書き込みデータの新たな組みを選択し、その書き込みデータの組みに対して新に誤り訂正用データを生成し、もって、新たな誤り訂正データグループを生成し、(e4) 該生成された新たな誤り訂正データグループを該

7

複数のドライブ内の空き領域に並列に書き込み、(e5) 該ステップ(e1)により読み出された書き込みデータが、該ステップ(e4)により再書き込みされたとき、該複数のドライブに保持されている、その再書き込みされる前のその書き込みデータを無効にするステップを有するデータ書き込み方法。

【請求項２１】上記ステップ(e1)は、いずれか一つの部分的に無効な誤り訂正データグループに含まれる有効なデータをすべて読み出すステップからなる請求項２０記載のデータ書き込み方法。

【請求項２２】上記ステップ(e3)から(e4)は、該ステップ(e1)または(e2)で読み出された書き込みデータが属する部分的に無効な誤り訂正データグループが特定状態にあるときに行なわれる請求項２１記載のデータ書き込み方法。

【請求項２３】該特定の状態は、部分的に無効な誤り訂正データグループに含まれた無効とされた書き込みデータの数が、予め定めた限界数以上である状態である請求項２２記載のデータ書き込み方法。

【請求項２４】データの書き込み領域をユーザの要求に従い確保するときに、その領域が動的アドレス変換をする領域か否かをユーザの指示により決定し、
その後上記一つまたは複数の上位装置から上記領域に書き込むべきデータが転送されたときに、上記書き込み領域が動的アドレス変換をする領域と決定されている場合に、上記ステップ(a)から(d)を実行する請求項１記載のデータ書き込み方法。

【請求項２５】複数のディスクドライブを有するディスクアレイシステムであって、
(a) それぞれ該複数のディスクドライブに新に書き込まれるべき所定の長さを有するデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定長を有するデータのいずれか一つからなり、一つまたは複数の上位装置から転送された複数の書き込みデータを保持するキャッシュメモリと、
(b) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正を用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成する回路と、
(c) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込む手段と、
(d) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にする手段とを有するもの。

【請求項２６】該書き込み手段は、該無効手段による無効を複数群の書き込みデータに対して実行した結果、該複数のドライブに記憶されていたいずれか一つの誤り訂正データグループに属する一群の書き込みデータがすべて無効となったとき、その一群の書き込みデータおよびその誤り訂正データグループに属する誤り訂正用データを保持していた、該複数のドライブ内の複数の領域を、空き領域として他の誤り訂正データグループの書き込みに使用する手段を有する請求項２５記載のディスクアレイシステム

8

【請求項２７】(e) 該複数のドライブの一つに障害が発生したとき、その一つのドライブに保持されていた複数のデータのうち、複数の有効なデータの各々を、該一つのドライブ以外のドライブに保持されている、そのデータが属する誤り訂正データグループに属する、複数の他の有効なあるいは無効な書き込みデータと誤り訂正用データとを用いて選択的に回復する手段をさらに有する請求項２５記載のディスクアレイシステム。

【請求項２８】(e) それぞれ少なくとも一つの有効な書き込みデータと少なくとも一つの無効な書き込みデータを含む複数の誤り訂正データグループ（部分的に無効な誤り訂正データグループ）に含まれる複数の有効な書き込みデータを詰め替える手段をさらに有し、その詰め替え手段は、(e1) 部分的に無効な誤り訂正データグループに含まれている一つまたは複数の有効な書き込みデータを、該複数のドライブから選択的に読み出し、読み出された一つまたは複数の有効な書き込みデータを該キャッシュメモリに書き込む手段と、(e2) 該書き込み手段(e1)による書き込みを他の部分的に無効な誤り訂正データグループに対して実行する手段と、(e3) 該キャッシュ記憶に保持された、複数の有効な書き込みデータから選択された該所定数に等しい数の有効な書き込みデータの新たな組みとその書き込みデータの組みに対して新に誤り訂正用データを生成し、もって、新たな誤り訂正データグループを生成する手段と、(e4) 該生成された新たな誤り訂正データグループを該複数のドライブ内の空き領域に並列に再度書き込みする手段と、(e5) 該手段(e3)(e4)を他の新たな誤り訂正データグループのために繰返す手段と、(e6) 複数の部分無効誤り訂正データグループであって、そこに含まれていた有効な書き込みデータが該ステップ(d4)(d5)により、すべて再書き込みされたものを保持していた複数の領域を、空き領域として他の誤り訂正データグループの書き込みに使用する手段とを有する請求項２５記載のディスクアレイシステム。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は高性能な入出力動作を可能とするディスクアレイ装置およびそのためのデータ方法。

【０００２】

【従来の技術】現在のコンピュータシステムにおいて

9

は、ＣＰＵ等の上位側が必要とするデータは２次記憶装置に格納され、ＣＰＵが必要とする時に応じ２次記憶装置に対してデータの書込み、読みだしを行っている。この２次記憶装置としては一般に不揮発な記憶媒体が使用され、代表的なものとして磁気ディスク装置、光ディスクなどがあげられる。以下、ディスク装置をドライブと呼ぶ。

【０００３】近年高度情報化に伴い、コンピュータシステムにおいて、２次記憶装置の高性能化が要求されてきた。その一つの解として、多数の比較的容量の小さなドライブにより構成されるディスクアレイが考えられている。

【０００４】D.Patterson,G.Gibson,and R.H.Kartz;A Case for Redundant Arrays of Inexpensive Disks(RAID)において、データを分割して並列に処理を行うディスクアレイ（レベル３）とデータを分散して、独立に扱うディスクアレイ（レベル５）について、その性能および信頼性の検討結果が報告されている。

【０００５】まず、データを分割して並列に処理を行うレベル３のディスクアレイについて説明する。ディスクアレイは多数の比較的容量の小さなドライブにより構成される。ＣＰＵから転送されてきた一つの書き込みデータを分割し、分割後の複数のデータからパリティデータを作成し、これらのデータを複数のドライブに並列に格納する。また、読み出す場合は逆に各々のドライブから分割されたデータを並列に読み込んできてそれらを結合してＣＰＵへ転送する。複数のデータと誤り訂正用のデータとの群をパリティグループと呼ばれることがある。本明細書でもこの用語を誤り訂正用データがパリティデータでない場合にも用いる。このパリティデータは分割後のデータを格納したドライブの中の１台に障害が発生し、データが読み出せなくなった場合、残りの正常なドライブ内のデータとパリティデータから、障害が発生したドライブ内のデータを復元するためのものである。ディスクアレイのような多数のドライブにより構成される装置では、部品点数が増加することにより、障害が発生する確率が高くなるため、信頼性の向上を図る目的で、このようにパリティデータを用意する。

【０００６】次に、データを分散して、独立に扱うレベル５のディスクアレイについて説明する。これは個々のデータを分割せずに独立に扱い、複数のデータからパリティデータを生成し、これらを多数の比較的容量の小さなドライブに分散して格納するものである。このパリティデータも先に述べたように、データを格納したドライブに障害が発生した場合、その障害ディスク内のデータを復元するためのものである。

【０００７】現在、一般に使用されている汎用大型コンピュータシステムの２次記憶装置では、他の読み出し／書込み要求に当該ドライブが使用されているため、そのドライブを使用できずに待たされることが多く発生し

10

た。このディスクアレイではデータを分散して格納してあるため、読み出し／書込み要求が増加してもディスクアレイの複数のドライブに分散して処理するため、読み出し／書込み要求がまたされることが減少する。

【０００８】これらのディスクアレイでは、現在一般に使用されている汎用大型コンピュータシステムと同様、２次記憶装置内では、個々のデータの格納場所（アドレス）は予め指定したアドレスに固定され、ＣＰＵから当該データへ読みだしまたは書込みする場合、この固定されたアドレスへアクセスすることになっている。

【０００９】データを分散して、独立に扱うレベル５タイプのディスクアレイでは、書込み時に新しくパリティデータを作成する際に大きな処理オーバヘッドが必要になる。以下それについて説明する。

【００１０】図１１は上記文献で示したD.Pattersonらが提案したＲＡＩＤに述べられている、データを分散して、独立に扱うレベル５タイプのディスクアレイにおける、書き込み時のパリティデータ作成方法を示す。この各アドレスにあるデータは１回の読み出し／書込み処理でアクセスされる単位で、個々のデータは独立である。また、ＲＡＩＤで述べられているアーキテクチャではデータに対するアドレスは固定されている。前述したようにこのようなシステムでは、信頼性を向上するためパリティデータを設定することが不可欠である。本システムでは各ドライブ内の同一アドレスのデータによりパリティデータが作成される。すなわち、ドライブ＃１から４までのアドレス（１．１）のデータによりパリティデータが作成され、パリティデータを格納するドライブの（１．１）に格納される。本システムでは読み出し／書込み処理は現在の汎用大型計算機システムと同様に各ドライブに対し当該データをアクセスする。このようなディスクアレイにおいて、例えばドライブ＃３のアドレス（２．２）にデータを書き込む場合、まず、ドライブ＃３の（２．２）の旧データと、ドライブ＃５の（２．２）の旧パリティデータを読みだし（ステップ１）、これらと書き込むデータとで排他的論理和をとり、新たなパリティデータを作成する（ステップ２）。

【００１１】パリティデータの作成完了後、書き込みデータをドライブ＃３の（２．２）に、新パリティデータをドライブ＃５の（２．２）に格納する（ステップ３）。

【００１２】図１２に示すように、レベル５のディスクアレイでは、データの格納されているドライブ、パリティデータの格納されているドライブから古いデータとパリティデータを読み出す時に、ディスクを平均１／２回転待ち、それから読みだしてパリティデータを作成し、この新しく作成したパリティデータとデータを格納するため、さらに一回転待たなければならないため、データを書き替える場合平均で１．５回転待たなければならない。ドライブにおいては１．５回転ディスクの回転を待

11

つということは非常に大きなオーバヘッドとなる。このような書込み時のオーバヘッドを削減するため、書込み先のアドレスを動的に変換する方法がストレージテクノロジコーポレーション（以下、ＳＴＫ）社から出願されているＰＣＴ国際公開公報WO　９１／１６７１１，ＷＯ　９１／２００７６に開示されている。

【００１３】この従来技術によれば、いずれかの、ＣＰＵが指定したデータが更新されるときには、そのデータを含む仮想トラックに属するデータ全体をキャッシュメモリに読み出し、その一部を更新データで更新した上で、更新後の、仮想トラックに属するデータがキャッシュから追い出されたときに、そのデータを物理トラックのセクタ単位に分割し、分割後のデータからあるいはそれらと他の書き込みデータとから新たなパリティグループを生成し、それをドライブ内の空き領域に書き込むようになっている。この際元の仮想トラックに属していてデータは無効にされる。パリティグループを構成するデータの長さは、ドライブの一つの物理トラックの容量を有するように定められている。適当なタイミングで、無効な書き込みデータを含む部分的に無効なシリンダから、有効なデータを収集して他の領域に書き込むことにより、そのシリンダを空き領域のみからなるシリンダにするようになっている。

【００１４】

【発明が解決しようとする課題】この方法によれば、パリティデータを構成するデータは、物理的トラックの長さを有するため、パリティグループを形成するまで複数のデータを保持するキャッシュの容量は大きくなる。

【００１５】本発明の目的は、更新用のデータを保持するより少ない容量のキャッシュメモリを使用して、動的にアドレス変換により更新データの書き込みを行いうるデータ書き込み方法およびそのための装置を提供することを目的とする。

【００１６】

【課題を解決するための手段】本発明では、複数のディスクドライブと、一つまたは複数の上位装置から転送された、それぞれ該複数のディスクドライブに新に書き込まれるべき所定の長さのデータあるいは該複数のディスクドライブに書き込み済みのデータを更新すべき該所定の長さを有するデータのいずれかからなる複数の書き込みデータを一時的に保持するキャッシュメモリとを有するディスクアレイシステムにおいて、(a) 該キャッシュメモリに保持された複数の書き込みデータから、誤り訂正用データグループを構成するためのデータとして、それぞれ該所定長を有する該所定数に等しい数の書き込みデータを取り出し、取り出された該所定数の書き込みデータから誤り訂正用のデータを生成し、(b) 該所定数の書き込みデータと該誤り訂正用のデータとを、一つの誤り訂正データグループとして、該複数のドライブの内の互いに異なるものに属する複数の空き領域に並列に書き込み、(c) 該所定数の書き込みデータが、該複数のディスクドライブに書き込み済みのデータに対する更新データを含む場合には、その書き込み済みのデータを無効にする。

12

【００１７】その後、適当なタイミングで、部分的に無効となったパリティグループ内の有効なデータを集めて、新たなパリティグループを生成し、空き領域に書き込み、元の部分的に無効なパリティグループ内の有効なデータをすべて無効とする。

【００１８】

【作用】書き替え前のデータ、パリティデータを読みださずにパリティデータを作成でき、書き替えのための旧データ、旧パリティデータの読み出しが不要となった。

【００１９】また、パリティグループを構成するデータの長さを上位装置から送られてくる一定のデータ長にしたので、パリティグループを生成するまでに書き込みデータを保持するためのキャッシュの容量が小さくてよい。

【００２０】

【実施例】

（実施例１）

（１）概要

以下本発明の一実施例を図１により説明する。

【００２１】本実施例はＣＰＵ１、アレイディスクコントローラ（以下ＡＤＣ）２，アレイディスクユニット（以下ＡＤＵ）３により構成される。ＡＤＵ３は複数の論理グループ１０により構成され，個々の論理グループ１０はm台のＳＣＳＩドライブ１２と、各々のＳＣＳＩドライブ１２とＡＤＣ２を接続するドライブパス９－１から４により構成される。なお，このＳＣＳＩドライブ１２の数は本発明の効果を得るには、特に制限は無い。この論理グループ１０は障害回復単位で、この論理グループ１０内のＳＣＳＩドライブ１２は、m－１個のデータとそれらから生成したパリティデータとからなる誤り訂正データグループ（以下では、簡単化のためにパリティグループと呼ぶ）を保持する。

【００２２】次にＡＤＣ２の内部構造について図１を用いて説明する。ＡＤＣ２はチャネルパスディレクタ５と２個のクラスタ１３とバッテリバックアップ等により不揮発化された半導体メモリであるキャッシュメモリ７により構成される。このキャッシュメモリ７にはデータとアドレス変換用テーブルが格納されている。このキャッシュメモリ７およびその中のアドレス変換用テーブルはＡＤＣ２内の全てのクラスタにおいて共有で使用される。クラスタ１３はＡＤＣ２内において独立に動作可能なパスの集合で、各クラスタ１３間においては電源、回路は全く独立となっている。クラスタ１３はチャネル、キャッシュメモリ７間のパスである、チャネルパス６と、キャッシュメモリ７、ＳＣＳＩドライブ１２間のパスであるドライブパス６－１から４が、それぞれ、２個

13

ずつで構成されている。それぞれのチャネルパス６－１から４とドライブパス８はキャッシュメモリ７を介して接続されている。ＣＰＵ１より発行されたコマンドは外部インターフェースパス４を通ってＡＤＣ２のチャネルパスディレクタ５に発行される。ＡＤＣ２は２個のクラスタ１３により構成され、それぞれのクラスタは２個のパスで構成されるため、ＡＤＣ２は合計４個のパスにより構成される。このことから、ＡＤＣ２ではＣＰＵ１からのコマンドを同時に４個まで受け付けることが可能である。そこで、ＣＰＵ１からコマンドが発行された場合ＡＤＣ２内のチャネルパスディレクタ５によりコマンドの受付が可能かどうか判断する。図２は図１のチャネルパスディレクタ５と１クラスタ１３－１内の内部構造を示した図である。図２に示すように、ＣＰＵ１からＡＤＣ２に送られてきたコマンドはインターフェースアダプタ１５により取り込まれ、マイクロプロセッサ（ＭＰ）２０はクラスタ内の外部インターフェースパス４の中で使用可能なパスがあるかを調べ、使用可能な外部インターフェースパス４がある場合はＭＰ２０はチャネルパススイッチ１６を切り換えてコマンドの受付け処理を行ない、受け付けられない場合は受付不可の応答をＣＰＵ１へ送る。

【００２３】レベル３のＲＡＩＤでは、基本的には、複数の書き込みデータから誤り訂正用のデータ、例えばパリティデータを生成し、上記複数の書き込みデータと誤り訂正用のデータとを、一つの誤り訂正データグループとして複数のドライブに分散して記憶される。

【００２４】誤り訂正用のデータには、パリティデータ以外のデータも使用可能である。しかし以下では、表現の簡単化のために、パリティデータ以外のデータを使用する場合を含めて、誤り訂正用のデータを単にパリティデータと呼び、誤り訂正データグループをパリティグループと呼ぶ。

【００２５】本実施例の概略的な動作は以下のとおりである。

【００２６】ＣＰＵから転送された書き込みデータを上記第１の領域に書き込む方法は、基本的には、従来のレベル５のＲＡＩＤによる方法に従う。

【００２７】すなわち、複数の書き込みデータをキャッシュメモリ７に保持し、保持された所定数のそれぞれ物理ドライブのレコード長に等しい長さの書き込みデータからパリティデータを生成し、これらのデータをパリティグループとして、いずれかの論理グループ１２内の複数のドライブに設けられた第１の領域内に分散して書き込む。

【００２８】書き込み済みのデータの読み出しをＣＰＵから要求されたときには、そのデータのみを選択的に読み出し、キャッシュメモリ７を介してＣＰＵに送る。

【００２９】書き込み済みのデータを更新するデータがＣＰＵから転送されたときには、その書き込み済みのデータとその書き込み済みのデータが属するパリティグループ内のパリティデータをキャッシュメモリ７に読み出し、これらの読み出されたデータと更新用のデータから、新たなパリティデータを生成し、この更新用のデータと新たなパリティデータでもって、古い書き込みデータと古いパリティデータを書き換える。従って、元のパリティグループ内のデータが部分的に更新されることになる。

14

【００３０】これに対して、ＣＰＵから転送された書き込みデータを上記第２の領域に書き込む方法は、本実施例で特徴的である。すなわち、ＣＰＵから転送され、キャッシュメモリ７に保持された書き込みデータは、いずれかの論理グループに書き込まれているデータを書き換える更新用のデータであるか否かにかかわらず、それぞれレコード長に等しい、該所定数分集められ、それらからパリティデータが生成され、新たなパリティデータグループとして、いずれかの論理グループ１２内の複数のドライブ内の設けた第２の領域内の空き領域に分散して記憶される。この結果、従来技術の問題のところで述べた、書き込みデータを更新するときのオーバヘッドは減少する。

【００３１】この複数の書き込まれたデータのうちのいずれか一つが、書き込み済みのデータを更新するデータであるときには、その書き込み済みのデータが無効にされる。この結果、その書き込み済みのデータが属していたパリティグループは、部分的に無効にされることになる。

【００３２】以上の書き込み処理を複数のパリティグループに対して行なうと、論理グループ内の空き領域が減少する。このため、本実施例では、部分的に無効となった、パリティグループ内の有効なデータをキャッシュメモリ７に集め、それらから新たなパリティグループを作り、空き領域に記憶する。この収集が完了した、部分的に無効なパリティグループは、空き領域として新たなパリティグループの書き込みにその後使用される。

【００３３】このように、第２の領域にＣＰＵから指定されたあるアドレスを有するデータを書き込んだ後、そのアドレスを有する別の書き込みデータがＣＰＵから転送された場合、後のデータは、前のデータが書き込まれた、ドライブ内位置とは異なる位置に記憶される。つまり、ＣＰＵが指定するアドレスを動的にドライブ内アドレスに変換する。従って、以下では、第２の領域を、動的アドレス変換をする領域とも呼ぶ。

【００３４】また、本実施例では、更新されるデータはドライブからは読み出されないで、無効にされる。このため、前述の国際公開公報に記載された動的変換とはことなり、データの更新を高速に行ない得る。

【００３５】（２）アドレス変換テーブル

本実施例ではＡＤＵ３を構成するＳＣＳＩドライブ１２はＳＣＳＩインターフェースのドライブを使用する。Ｃ

15

PU1をIBMシステム9000シリーズのような大型汎用計算機とした場合、CPU1からはIBMオペレーティングシステム（OS）で動作可能なチャネルインターフェースのコマンド体系にのっとってコマンドが発行される。そこで、SCSIドライブ12をSCSIドライブを使用した場合、CPU1からのコマンドを、SCSIインターフェースのコマンド体系にのっとったコマンドに変換する必要が生じる。この変換はコマンドのプロトコル変換と、アドレス変換に大きく分けられる。

【0036】以下にアドレス変換用のテーブルについて説明する。

【0037】以下では、CPU1から転送されるデータ長はドライブの1セクタ長に等しいか、又はそのn倍（nは1より大きい整数）とする。セクタ長のn倍の書き込みデータ（レコード）を本実施例では、セクタ長のm倍（mはnより小さい整数）に等しい長さの複数のデータ（ブロック）に分割して処理する。しかし、以下では、簡単化のために、CPU1から転送されたデータのレコード長が常に一定な場合についてのみ説明する。CPU1から指定されるアドレスは、図13に示すようにデータが格納されているトラックが所属するシリンダの位置とそのシリンダ内において当該データが格納されているトラックを決定するヘッドアドレスと、そのトラック内のレコードの位置を特定する。具体的には要求データが格納されている当該ドライブの番号（ドライブ番号）と当該ドライブ内のシリンダ番号であるシリンダアドレス（CC）とシリンダにおいてトラックを選択するヘッドの番号であるヘッドアドレス（HH）とレコードアドレス（R）からなるCCHHRである。従来のCKDフォーマット対応の磁気ディスクサブシステム（IBM3990－3390）ではこのアドレスに従ってドライブへアクセスすれば良い。しかし、本実施例では複数のSCSIドライブ12により従来のCKDフォーマット対応の磁気ディスクサブシステムを論理的にエミュレートする。つまり、ADC2は複数のSCSIドライブ12が、従来のCKDフォーマット対応の磁気ディスクサブシステムで使用されているドライブ1台に相当するようにCPU1にみせかける。このため、CPU1から指定してきたアドレス（CCHHR）をSCSIドライブのアドレスにMP20が変換する。このアドレス変換には図3に示すようなアドレス変換用のテーブル70（以下アドレステーブルとする）が使用される。ADC2内のキャッシュメモリ7には、その内部の適当な領域にこのアドレステーブル70が格納されている。本実施例では、CPU1が指定してくるドライブはCKDフォーマット対応の単体ドライブである。しかし、本発明ではCPU1は単体と認識しているドライブが、実際は複数のSCSIドライブにより構成されるため、論理的なドライブとして定義される。このため、ADC2のMP20はCPU1より指定してきたCPU指定アドレス71（これはドライブ番号74とCCHHR75からなる）をSCSIドライブ12に対するSCSIドライブアドレス72（これはSCSIドライブ番号77とそのSCSIドライブ内のアドレス（以下SCSI内Addrとする）78からなる）に変換する。アドレステーブル70はCPU1が指定するCPU指定アドレス71とそれに対応する、実際にデータが格納されているSCSIドライブ12内のアドレス（SCSIドライブアドレス）72と、そのデータに対応したパリティデータの格納されているパリティドライブアドレス73とキャッシュメモリ7内のアドレス（キャッシュアドレス）81と、キャッシュメモリ7内に当該データが存在するかどうかのキャッシュフラグ82が格納される。このキャッシュフラグ82はオン（1）の場合キャッシュメモリ7内にデータが存在していることを示し、オフ（0）の場合はキャッシュメモリ7内にデータが存在していないことを示す。

16

【0038】また、アドレステーブル70には動的アドレス変換を行なう場合の図4に示す動的アドレス変換テーブル90へのポインタ（DMポインタ）76が格納される。SCSIドライブアドレス72はデータの格納されているSCSIドライブの番号（SCSIドライブ番号）77とSCSI内Addr78（これはSCSIドライブ内のシリンダアドレス、ヘッドアドレス、レコード番号、セクタ数からなる）により構成されており、パリティドライブアドレス73にはそのデータに対応したパリティデータの格納されているSCSIドライブの番号（パリティドライブ番号）79とパリティデータの格納されているSCSIドライブ内のアドレス（以下パリティデータ内Addrとする）80（これはパリティドライブ内のシリンダアドレス、ヘッドアドレス、レコード番号、セクタ数からなる）が格納される。なお、アクセスフラグ84は、実施例2で使用するものである。

【0039】キャッシュアドレス81はDMポインタ76により対応するSCSIドライブアドレス72により決定したデータが、キャッシュメモリ7内に有る場合は、このデータのキャッシュメモリ7内のアドレスを登録しておく。また、この様にキャッシュメモリ7内に当該データが格納されている場合は、アドレステーブル70と同様にキャッシュフラグ82をオン（1）とする。無効フラグはDMポインタ76に対応するSCSIドライブアドレス72により決定したデータが有効であるか無効であるかを示すフラグで、無効である場合はオン（1）が登録される。ドライブフラグはドライブ内に書き込まれているかどうかを示すフラグで、オン（1）の場合はこのデータはドライブに書き込まれており、オフ（0）の場合は、まだドライブには書き込まれていないことを示す。

【0040】なお、アドレステーブル70と動的アドレス変換テーブル90はシステムの電源をオンした時に、

17

MP20により論理グループ10内のある特定のSCSIドライブ12から、キャッシュメモリ7にCPU1の関与なくして自動的に読み込まれる。一方，電源をオフする時は、MP20によりキャッシュメモリ7内のアドレステーブル70を、読み込んできたSCSIドライブ12内の所定の場所にCPU1の関与なくして自動的に格納する。

【0041】（3）動的アドレス変換を行う領域とその他の領域の確保

以下、本実施例の処理を詳細に説明する。

【0042】まず、ユーザは、ディスクアレイを使用する前に動的アドレス変換を行なうデータを格納する領域と、通常のレベル5で処理する領域をCPU1を介して設定する。

【0043】本実施例による動的にアドレス変換を行なうと、空き領域を予め用意しておかなければならず、また、処理中に空き領域が無くなった場合は詰め変えを行い、空き領域を確保しなければならない。このため、本実施例では非常にランダム性の高いアクセスが生じるデータについては詰め変えが頻繁に発生するため、なるべく動的にアドレス変換を行なわない。そこで、ユーザは動的にアドレス変換を行うことで性能向上が図れるシーケンシャルデータの量を予め調査し、動的アドレス変換を行なう領域を設定する。設定方法は、初期設定時にユーザは確保したい領域をCPU指定アドレスの特定の範囲の形でADC2に申請する。ADC2ではMP20が図3、4に示すように、このユーザが要求する動的アドレス変換を行なう領域の大きさに見合う領域を、SCSIドライブ12内に確保し、このSCSIドライブ12内に確保した領域のアドレス（SCSI内Addr78）を動的アドレス変換テーブル90に登録する。MP1　20はアドレステーブル70の一つの行と動的アドレス変換テーブル90の一つの行との間にリンクをはるために、それぞれに同じ値のDMポインタ76をセットする。上述のCPU指定範囲に属するCPU指定アドレスを特定アドレスと呼ぶ。後にCPU1からデータの書き込みを要求するときに、動的アドレス変換を行う場合、前述のCPU指定範囲に属する特定アドレスを書き込み先として指定したとき、このCPU指定アドレスは、アドレステーブル70内のDMポインタ76が設定されている一つの行に登録される。上記CPU指定範囲に属さないCPU指定アドレスが書き込み要求で指定されたときには、アドレステーブル70内の、DMポインタ76が設定されていない一つの行にこのCPU指定アドレスが登録される。また、動的アドレス変換はシーケンシャルデータのみでなく、データ圧縮を行なうデータに関しても動的アドレス変換を行なう領域に格納する。

【0044】動的アドレス変換を行う領域は、各論理グループ10を構成する複数のドライブ内の同じアドレス範囲の領域に分散して構成する。それぞれの領域は、それぞれセクタ単位の大きさを有する複数の領域から構成してもよく、あるいはシリンダ単位の大きさを有する複数の領域から構成してもよい。この様に領域を設定された後、実際の処理としては、CPU1が指定してきたアドレス（CPU指定アドレス71）をMP20がキャッシュメモリ7内のアドレステーブル70によりアドレス変換する際、CPU指定アドレス71が動的アドレス変換を行なう領域のアドレス（特定アドレス）かどうかをDMポインタ76を参照し、特定アドレスのデータであれば、動的アドレス変換を行なうように処理し、特定アドレスでないアドレスのデータであれば通常のレベル5のように処理する。

18

【0045】（4）データ書き込み動作の概要

以上のような動的アドレス変換を行なう領域とレベル5の領域を設定した後、以下のようなデータの書き込み処理を行なう。

【0046】CPU1から書き込み命令が発行されたとする。まず、ADC2のいずれかのMP20はCPU1からCPU指定のドライブ番号と、書込みデータとCPU指定アドレス71を指定する書き込みコマンドを受け取った後、そのMP20が所属するクラスタ13内の各チャネルパス6において処理可能かどうかを調べ、可能な場合は処理可能だという応答をCPU1へ返す。CPU1では処理可能だという応答を受け取った後にADC2へデータを転送する。この時、ADC2ではMP20の指示によりチャネルパスディレクタ5において、チャネルパススイッチ16が当該外部インターフェースパス4とインタフェースアダプタ15を当該チャネルパス6と接続しCPU1とADC2間の接続を確立する。CPU1とADC2間の接続を確立後CPU1からのデータ転送を受け付ける。CPU1から転送されてきたデータはMP20の指示により、チャネルインタフェース21によりプロトコル変換を行ない、外部インターフェースパス4での転送速度からADC2内での処理速度に速度調整する。チャネルインタフェース21におけるプロトコル変換および速度制御の完了後、データはDCC22によるデータ転送制御を受け、圧縮回路27に転送される。

【0047】データ圧縮を行なうデータについては圧縮回路27に転送されてきたデータはMP20の指示によりデータ圧縮される。

【0048】圧縮回路27により圧縮されたデータは、チヤネルアダプタ24に転送され、そのアダプタ24によりキャッシュメモリ7内に格納される。データ圧縮を行なわないデータはDCCから圧縮回路27内のスルーパスを通りチャネルアダプタ24に転送され、圧縮データと同様にキャッシュメモリ7内に格納される。書き込みデータを格納するキャッシュメモリ7はバッテリ等により不揮発化されている方が好ましい。この様にキャッシュメモリ7にデータを格納したのをMP20が確認し

たら、ＭＰ２０は書込み処理の完了報告をＣＰＵ１に対し報告する。

【００４９】このような動作を繰り返して、複数の書き込みデータが、キャッシュメモリ７に書き込まれる。この際、その書き込みデータに対して指定されたアドレスが前述のＣＰＵ指定アドレスの特定の範囲に属するものか否かをアドレステーブル７０に基づいて判断され、その結果を用いて、これらの書き込みデータは、動的書き込み領域に書き込むべきデータのグループとそうでないデータのグループに分かれてキャッシュメモリ７内に記憶される。これらの書き込みデータは、それぞれのグループ毎に処理される。

【００５０】動的アドレス変換をしない領域に書き込まれるべきデータについては以下のように処理される。

【００５１】まず、それぞれの書き込みデータに対してＣＰＵが指定したアドレスが、すでにアドレス変換テーブルに登録されているか否か、すなわち、それぞれの書き込みデータが、すでに書き込み済みのデータを更新するための書き込みデータであるかあるいはそのアドレスのデータが初めて書き込まれるものかが判断される。

【００５２】書き込み済みのデータを更新しないデータは、それらが所定数に達すると、ＭＰ２０の指示により、パリティデータ発生回路（ＰＧ）３６が、それらのデータからパリティデータグループを生成し、ＭＰ２０は、それらの書き込みデータおよびパリティデータを、書き込むべき空き領域を決定する。

【００５３】すなわち、いずれかの論理グループ１０内の所定数のドライブ内の、上記領域に属する、互いにドライブ内アドレスが等しい空き領域をアドレステーブル７０を参照して選択する。その後、アドレステーブル７０にこれらの書き込みデータのＣＰＵ指定アドレスを登録し、これらの書き込みデータに割り当てられた記録領域のドライブ番号７７及びドライブ内アドレス７８、およびパリティデータに関する同様のアドレス７９、８０を登録し、キャッシュアドレス８１を登録した上で、キャッシュフラグ８２をオン（１）にセットする。

【００５４】しかる後、これらのパリティグループを、上で割り当てられた空き領域に記憶する。

【００５５】動的アドレス変換をしない領域に書き込まれるべきデータが、書き込み済みのデータを更新するデータであると判断された場合には、以下に詳しく説明するように、その書き込み済みデータとその書き込み済みのデータが属する元のパリティグループ内の元のパリティデータが読み出され、それらのデータと更新用の書き込みデータから新たなパリティデータが生成され、その更新用のデータとその新たなパリティデータでもって、元の書き込み済みのデータおよび元のパリティデータを書き換える。従って、元のパリティグループが書き換えられるのであって、新たなパリティグループは生成されない。

【００５６】以下この場合の書き込み動作を図１１を参照して説明する。

【００５７】

（５）非動的アドレス変換をする領域内のデータの更新

まずＭＰ１　２０がドライブインタフェース２８に対し書き込み済みのデータとパリティデータを読みだすように指示を出す。具体的には図３においてＣＰＵ指定アドレス７１がＤｒｉｖｅ＃２，ＡＤＲ３の位置にあるデータを書きかえる場合、ＭＰ２０はアドレステーブル７０においてＣＰＵ指定アドレス７１（Ｄｒｉｖｅ＃２，ＡＤＲ３）に対応する各項目を調べる。図３においてＭＰ２０はこのＣＰＵ指定アドレスはすでに登録されているがＤＭポインタ７６がそのアドレスに対して登録されていないため、このデータはレベル５で処理するデータと判断し、ＳＣＳＩドライブアドレス７２とパリティドライブアドレス７３から書き込み先のＳＣＳＩドライブ番号７７とＳＣＳＩ内Ａｄｄｒ７８とパリティドライブ番号７９とパリティデータ内Ａｄｄｒ８０を認識する。ＭＰ２０はこの変換後のアドレスによりドライブインタフェース２８に対し書き込み先のアドレスにすでに格納されている書き込み前のデータおよびパリティデータの格納されているＳＣＳＩドライブ１２に対し、当該データおよび当該パリティデータの読み出し要求を発行するように指示を出す。ドライブインタフェース２８からこの読み出し要求を受けた２つのＳＣＳＩドライブ１２においては当該ＳＣＳＩ内Ａｄｄｒ７８、およびパリティデータ内Ａｄｄｒ８０に対しアクセスし、書き込み済みのデータおよびパリティデータをそれぞれの当該ＳＣＳＩドライブからキャッシュメモリ７に転送する。パリティデータ生成回路３６はこれらの書込み済みのデータとパリティデータと、新たな書き込みデータとからＭＰ２０の指示で、更新後の新しいパリティデータを作成する。新たな書き込みデータと新たなパリティデータをそれぞれ更新前の書き込みデータおよびパリティデータが格納されていたアドレスに書き込む。

【００５８】（６）動的にアドレス変換をする領域内のデータの書き込みと更新

このように、動的アドレス変換をしない領域に書き込まれるべきデータは、そのデータがすでの書き込み済みのデータを更新するデータであるか否かにより異なる処理を受ける。

【００５９】一方、動的アドレス変換をする領域に書き込まれるべきデータについては、以下のように処理される。

【００６０】そのようなデータは、書き込み済みのデータを更新するデータであるか否かに依らずに，所定数ずつまとめて、新たなパリティグループとして論理グループ内の空き領域に記憶される。

【００６１】すなわち、そのような書き込みデータが所定数に達した場合には、ＭＰ２０の指示により、パリテ

ィデータ生成回路３６がそれらのデータからパリティデータを生成し、それらの書き込みデータと生成されたパリティデータとから、新たなパリティグループを生成する。ＭＰ２０は、それらのパリティグループを書き込むべき空き領域を決定する。先の、動的アドレス変換をしない場合とことなり、空き領域は、動的アドレス変換用の領域に属する領域から選択する。

【００６２】すなわち、いずれかの論理グループ１０内の所定数のドライブの互いにドライブ内アドレスが等しい空き領域を動的アドレステーブル９０を参照して選択する。その後、アドレステーブル７０にこれらの書き込みデータのＣＰＵ指定アドレスを登録し、さらに、動的アドレステーブル９０に、これらの書き込みデータに割り当てられた記録領域のドライブ番号７７及びドライブ内アドレス７８、およびパリティデータに関する同様のアドレス７９、８０を登録し、キャッシュアドレス８１を登録した上で、キャッシュフラグ８２をオン（１）にセットし、かつ、無効フラグ９１、ドライブフラグ８３は０のままにする。

【００６３】しかる後、これらのパリティグループを、上で割り当てられた空き領域に記憶する。

【００６４】この新たなパリティグループを構成する書き込みデータのいずれもが、書き込み済みデータを書き換えるデータでないときには、書き込み動作は以上で終了するが、それらの書き込みデータのいずれかが、すでに書き込み済みのデータを更新する場合、その書き込み済みのデータを無効にする。従って、元のパリティグループは、部分的に無効なデータを含むことになる。

【００６５】以下、動的アドレス変換をする領域へのデータの書き込みの詳細を図６を参照して説明する。

【００６６】例えばＳＣＳＩドライブＳＤ＃１の内に書き込み済みデータ＃１、次にＳＤ＃２書き込み済みデータ＃２，次にＳＤ＃３のデータ＃３，次にＳＤ＃４に書き込み済みのデータ＃４に対して順に書き込み命令がＣＰＵより発行された場合を例にして説明する。

【００６７】ＭＰ２０はこれらのデータをキャッシュメモリ７に書き込む。このとき、キャッシュ内のアドレステーブル７０，９０にそれぞれのデータについての情報を登録する。なお、更新前の書き込みデータ＃１～＃４がキャッシュメモリ７内に保持されている場合には、キャッシュメモリ７内のそれらの旧データを書き替えた後、以下のことを行う。

【００６８】論理グループ１０を５台のＳＣＳＩドライブ１２で構成すると仮定する。ＭＰ２０は、キャッシュメモリ７に４個の書き込みデータが揃うとパリティデータの作成を行う。

【００６９】すなわち、ＭＰ２０は、動的アドレス変換テーブル９０において、無効フラグ９１がオフ（０）でキャッシュフラグ８２がオン（１）でドライブフラグ８３がオフ（０）のデータを探す。ＭＰ２０はこの条件を

満足するデータの数を調べその数が４に達し次第、その４つの書き込みデータについて、パリティデータ生成回路（ＰＧ）３６にパリティデータの作成を指示する。ＭＰ２０は、これら４つのデータと新たに生成したパリティデータとを、一つの新たなパリティグループとして論理ドライブに並列に書き込む。このために、ＭＰ２０はまず書き込み先のアドレスを探す。つまり、各ＳＣＳＩドライブ１２においてデータ＃１から４とパリティデータが全て書込み可能なスペースをさがす。具体的には、ＭＰ２０は動的アドレス変換テーブル９０において、論理グループを構成するＳＣＳＩドライブ１２に対し、同一ＳＣＳＩ内Ａｄｄｒ７８の無効フラグ９１がオンになっている領域を探す。この場合、対応するパリティデータが有効か否かは問わない。

【００７０】以下ではデータ＃１から４のＣＰＵ指定アドレス７１が、図４に示す動的アドレス変換テーブル９０においてデータ＃１がＤＭ　ａ－２，データ＃２がＤＭｂ－３、データ＃３がＤＭ　ｃ－１，データ＃４がＤＭ　ｄ－３に対応するとする。

【００７１】図４の場合ＳＣＳＩドライブ＃１から５が論理グループを構成するとしてドライブＳＤ＃１から４のアドレスＳＡＤＲ４の領域は全て無効フラグ９１がオンになっており、データ＃１から４を書き込める空き領域と判断する。

【００７２】そこで、ユーザがデータ＃１の書き込み先に指定してきたＣＰＵ指定アドレス７１（特定アドレス）に対し、ＭＰ１　２０はＤＭポインタ７６により動的アドレス変換テーブル９０のＤＭ　ａ－２の項目で、ドライブＳＤ＃１のＳＡＤＲ２に書き込む様に変換されるが、動的アドレス変換を行なうため、書き込みスペースの判断後、新たな書き込み先と決定した、ＳＣＳＩドライブＳＤ＃１内のＳＡＤＲ４に書き込むことを決定する。同様に、ＭＰ２０はドライブＳＤ＃２のアドレスＳＡＤＲ３のデータ＃２、ＳＤ＃３のＳＣＳＩドライブ１２のＳＡＤＲ１のデータ＃３、ドライブＳＤ＃４のＳＡＤＲ３のデータ＃４をそれぞれ更新する書き込みデータ＃２，＃３，＃４も、ＳＣＳＩドライブＳＤ＃２～ＳＤ＃４のアドレスＳＡＤＲ４に書き込むことを決定する。

【００７３】この様に、ＭＰ２０が書き込み先のアドレスを決定した後、ＭＰ２０が上記書き込み可能なスペースのＳＣＳＩ内Ａｄｄｒ７８（ＳＡＤＲ４）に対し、ドライブインタフェース２８に、各ＳＣＳＩドライブＳＤ＃１～ＳＤ＃５への書込み要求を発行するように指示する。ドライブインタフェース２８ではＳＣＳＩの書込み処理手順に従って、当該ＳＣＳＩドライブに対し書込みコマンドとＳＣＳＩ内Ａｄｄｒ７８（ＳＡＤＲ４）をドライブユニットバス９を介して発行する。ドライブインタフェース２８から書込みコマンドを発行された各ＳＣＳＩドライブ１２においては指示ＳＣＳＩ内Ａｄｄｒ７８（ＳＡＤＲ４）へシーク、回転待ちのアクセス処理を

行なう。当該ＳＣＳＩドライブ１２におけるアクセス処理が完了した後、キャッシュアダプタ１４はキャッシュメモリ７からデータを読み出してドライブインタフェース２８へ転送する。ドライブインタフェース２８は転送されてきたデータをドライブユニットパス９を介して該当するＳＣＳＩドライブ１２へ転送する。データの当該ＳＣＳＩドライブ１２のＳＣＳＩ内Ａｄｄｒ７８（ＳＡＤＲ４）への書込みが完了すると、当該ＳＣＳＩドライブ１２はドライブインタフェース２８に完了報告を行ない、ドライブインタフェース２８がこの完了報告を受け取ったことを、ＭＰ２０に報告する。この時、ＭＰ２０は、この書込みデータをキャッシュメモリ７上に残さない場合は、この報告を元にアドレステーブル７０のキャッシュフラグ８２をオフにする。パリティデータに関してもデータと同様にパリティドライブアドレス７３に従って書き込まれる。また、この様に書込みアドレスを動的に変換した場合は、ＳＣＳＩドライブ１２からの完了報告をドライブインタフェース２８が受け取った際に、アドレステーブル７０のユーザが書き込み先として指定したＣＰＵ指定アドレス７１に対するＤＭポインタ７６の値を、動的アドレス変換テーブル９０における書き込み後のＳＣＳＩドライブアドレス７２に対応するＤＭポインタ７６の値に変更する。この時、同時に動的アドレス変換テーブル９０の各項目も変更する。具体的には図３のアドレステーブル７０において、ＣＰＵ１がデータ＃１をＤｒｉｖｅ＃１のＡＤＲ１に書き込むように指示した場合、このデータ＃１を上記のように動的アドレス変換した場合、書き込み前はＤＭポインタ７６はＤＭ　ａ－２となっているため、動的アドレス変換テーブル９０において、ＳＣＳＩドライブＳＤ＃１のアドレスＳＡＤＲ２が対応していた。動的アドレス変換後により、データ＃１の書き込み後は、ＭＰ１　２０はアドレステーブル７０のＤｒｉｖｅ＃１，ＡＤＲ１のＤＭポインタ７６をＤＭ　ａ－２からＤＭ　ａ－４に変更する。これと同時にＭＰ２０は動的アドレス変換テーブル９０においてＤＭ　ａ－４の無効フラグ９１をオフ（０）、ドライブフラグ８３をオン（１）にし、書き込みデータをキャッシュ７内に残すときは、キャッシュアドレス８１にアドレスを登録しキャッシュフラグ８２をオン（１）のままとする。将来，ユーザがこのデータに読み出し要求を発行した場合は前のアドレスにはアクセスせず新しく格納されたアドレスにアクセスすることになる。

【００７４】また、同時に動的アドレス変換テーブル９０において、書き込み前のＤＭ　ａ－２の無効フラグ９１をオン（１）とし、キャッシュフラグ８２、ドライブフラグ８３をオフ（０）とする。この領域は後の別な書き込みのスペースとして使用される。

【００７５】以上のようなドライブへの書き込み動作は、論理グループ１０を構成する各ＳＣＳＩドライブ１２において並列に行なう。以下では以上の書き込み動作を一括書き込みとも呼ぶ。

【００７６】この様に動的アドレス変換をする領域に対しては、書き込みデータが書き込み済みのデータを更新するか否かに関係なく、書き込みデータをキャッシュメモリ７に溜め、溜められた所定数の書き込みデータによりパリティデータを作成し、書き込みデータとパリティデータを複数のＳＣＳＩドライブ１２へ並列に書き込む。従来のレベル５のようなユーザが指定した書き込み先のアドレスにすでに格納されている、書き込み前のデータとパリティデータに対する読みだしは行なわない。このため書き込み時のオーバヘッドを削減することが可能となる。

【００７７】しかし、書き込みデータについて動的にアドレスを変更し格納していくと、無効フラグ９１がオン（１）になっている領域が分散されるため、データ格納効率（ＳＣＳＩドライブに格納可能なデータ容量に対する、実際に格納するデータ容量の割合）が低下し、さらにこの様な状態が進行すると、動的アドレス変換による一括書き込みを行なう空き領域が無くなってしまう。これに対しては、後述するデータのつめ替えで対処する。

【００７８】なお、本実施例では同じドライブ内アドレスを有する複数の領域にデータを並列に格納するため、論理グループ１０内の各ＳＣＳＩドライブ１２の回転を全て同期させる方が望ましい。

【００７９】（７）障害回復

いくつかの無効データが存在する場合において、論理グループ１０内のＳＣＳＩドライブに障害が発生した場合、障害ＳＣＳＩドライブを除いた残りのＳＣＳＩドライブ１２内のデータおよびパリティデータにより障害ＳＣＳＩドライブ内のデータを復元することが可能となる。このような障害回復を行なう場合、各ＳＣＳＩドライブ１２内のキャッシュメモリ７内の対応する無効フラグ９１がオンになっているデータも有効なデータとして読みだして障害データの復元に使用する。無効フラグ９１がオンになったデータは更新前の古いデータであるが、それが属していたパリティグループが生成されたときの値を有している。したがって、障害ドライブのデータの回復には使用できる。このため、本実施例では、動的アドレス変換領域に書き込み済みのデータが更新された場合、そのデータに対して無効フラグをつけるが、データは消去しないのは、そのためである。

【００８０】

（８）動的アドレス変換領域と非動的アドレス変換領域

まず、データ格納領域について説明する。本実施例ではＳＣＳＩドライブ１２内の動的アドレス変換を行なう領域に実際に格納するデータの量は、動的アドレス変換を行なう領域に格納可能な最大容量の１／Ｖとする。これは、書き込みデータのアドレスを動的に変換して一括書き込みを行なうため、一括書き込みを行なう領域（書き替え領域）が必要となるからである。つまり、ＳＣＳＩ

ドライブ１２の動的アドレス変換を行なう領域に書込み可能な最大容量のデータを格納してしまうと、後に書き込み要求が発行された場合、一括書き込みを行なおうとしても書き込む場所が無くなってしまうためである。大型汎用計算機では、ユーザは予め自分の使用可能な領域（容量）を確保しておく。ユーザがデータの読みだし、書き込みを行なう場合は、この容量に見合う領域内で処理しなければならない。もし、この使用可能な領域（容量）を越えた場合は新たに領域（容量）を確保しなければならない。そこで、先に示したようにユーザがＣＰＵ１により動的アドレス変換を行なう領域を設定する際、予め後の書き込み要求のために書き替え領域を見込んで、領域を確保しておく（（Ｖ－１）／Ｖ）。

【００８１】この予め確保しておく書き替え領域の割合は、動的アドレス変換をシーケンシャルアクセスされるデータのみで行なうのではなく、一部ランダムアクセスされるデータについても行なう場合、ランダムデータに関し、ランダムなアドレスにアクセスする読み出し／書込み要求が多い場合は１／Ｖの値を小さくし、比較的同じようなアドレスにアクセスする読み出し／書込み要求が多い場合は１／Ｖの値を大きくするようにＶの値を変えることが望ましい。さらに、前者の場合、キャッシュメモリ７から早くＳＣＳＩドライブにデータを移し、後者の場合は逆にデータがキャッシュメモリ７内に長く滞留させることが望ましい。１／Ｖを変更する方法には２つの方法がある。一つは、ユーザが予め格納するデータについて，ランダムなアクセスの量を認識し，格納するデータの割合（１／Ｖ）を決める。ユーザは自分のデータの特性から判断し、自分が確保した領域（容量）の中で、実際に格納するデータの割合を決めるのである。例えば、同じデータに対して次々と書き換えを行なうアクセスの割合が、全アクセスに対して約３０％程度なら、１／Ｖ＝１／２程度に設定し、予めユーザが確保した領域（容量）の１／２のみユーザはデータを格納しておく。もう一つは、ユーザが予め領域（容量）を確保するのではなく、ＡＤＣ２内のＭＰ２０が一定時間アクセスされたアドレスの履歴を記録しておき，そのアドレス履歴データからランダムなアクセスの割合を求め、ＡＤＣ２のＭＰ２０が格納するデータの割合（１／Ｖ）を自動的に割り当てる。例えばＳＣＳＩドライブ１２に格納可能なデータ量の４／５のデータが格納されており、このデータに対してランダムなアクセスの割合が増加した場合、別の論理グループに随時データを移し、ＭＰ１　２０は１／Ｖを小さくする。

（９）部分無効パリティグループの詰め変え

しかし、このように予め書き替え領域を確保しておいても、１ＳＣＳＩドライブ１２内に格納できるデータの容量は有限であるため、一括書き込みを繰り返した場合、書き替え領域が無くなる。そこで、ＳＣＳＩドライブ１２内の有効なデータをまとめ、新たに書き替え領域を作成する必要が生じる。この方法について次に説明する。

【００８２】図７は各ＳＣＳＩドライブ１２内の同一ＳＣＳＩ内Ａｄｄｒ７８のデータを抜き出したものである。論理グループ１０内における各ＳＣＳＩドライブ１２の同一ＳＣＳＩ内Ａｄｄｒ７８（同一行）のデータつまり、同一のパリティグループを形成すべきデータにより、パリティデータが作成される。例えばＳＡＤＲ　ａのアドレスを考えると、ＳＣＳＩドライブ＃１のデータ１、ＳＣＳＩドライブ＃２のデータ＃５、ＳＣＳＩドライブ＃３のデータ＃９、ＳＣＳＩドライブ＃４のデータ＃１３によりパリティデータ＃１が作成される。図７ではＳＡＤＲ　ａのデータグループ内において無効フラグ９１がオフのデータはデータ＃９のみである。前述したように無効とされたデータ＃１，＃５，＃１３は障害発生時の障害ディスク内のデータ復元用データとして残されている。ＳＡＤＲ　ｂについては無効フラグ９１がオフのデータはデータ＃１０，＃１４の２個で、ＳＡＤＲ　ｃではデータ＃３，＃７，＃１５の３個、ＳＡＤＲ　ｄはデータ＃４，＃８，＃１２，＃１６と全て無効フラグ９１がオフとなっている。ＭＰ２０はキャッシュメモリ７内のアドレステーブル７０および動的アドレス変換テーブル９０を常に監視しており、動的アドレス変換テーブル９０内のＳＣＳＩドライブアドレス７２と無効フラグ９１をみて，同一パリティグループに属する、すなわち、同一ＳＣＳＩ内Ａｄｄｒ７８に対応した、無効フラグ９１のオン（１）となっている領域の数を認識しこの領域の数が予め設定しておいた数より多くなると、詰め変えを行う。例えば、同一ＳＣＳＩ内Ａｄｄｒ７８に対し無効フラグ９１がオン（１）になった領域の数が３個より多くなると詰め変えを行うように予め設定しておいた場合、図７のＳＡＤＲ　ａを図４のＳＡＤＲ３とすると、このように同一ＳＣＳＩ内Ａｄｄｒ７８に対し無効フラグ９１が３個オン（１）になると、ＭＰ２０は詰め変えを行う。また、この詰め変え処理はユーザの指示で行なうことも可能とする。

【００８３】具体的には、ＭＰ２０は、その部分無効パリティグループ内の有効なデータ、つまり、今の例では、図４のドライブＳＤ＃３のアドレスＳＡＤＲ３のデータに対し、擬似的な読みだし要求を発行し、このデータをキャッシュメモリ７に擬似的に読みだしキャッシュメモリ７内に溜め、ＣＰＵ１から転送された他の書き込みデータまたは同様に他の行から擬似的に読みだされたデータととからなる４つのデータを組にし、それらから新たにパリティデータを作成し、動的アドレス変換における書き込みと同時に書き込み用空きスペースを探し、それらの領域に並列に格納する。なお、データ＃９がキャッシュメモリ７に読み込まれた段階で、ＭＰ２０はキャッシュメモリ７内のアドレステーブル７０内のキャッシュアドレス８１を変更しキャッシュフラグ８２をオンにする。また、書き替え領域への格納後は、動的アドレ

27

ス変換の書き込みと同様にアドレステーブル７０および動的アドレス変換テーブル９０の各項目を変更する。

【００８４】また、別の詰め変え方法としては、上述したように部分無効パリティグループが詰め変え条件を満すとすぐ行うというのでなく、そのようなデータに読み出し要求がＣＰＵ１よりその後発行され、ＳＣＳＩドライブ１２から読み出された時に行う方法がある。すなわち、上位装置にそのデータを転送すると同時にキャッシュメモリ７内に溜め、書き込みデータまたは同様に他の行から吸い上げられたデータと新たにパリティデータを作成し、これらのデータパリティデータをパリティデータを格納するＳＣＳＩドライブ１２のそれぞれの書き替え領域に格納する。

【００８５】詰め替えを起動する契機としてはさらに次のものでもよい。

【００８６】すなわち、部分的に無効となったパリティグループの数が、予め決めた限界数を越えたときに行なう。この際、無効となったデータの数が一つでも含んでいるようなパリティグループの数をカウントして、そのカウント値が所定数を越えたときに、詰め替えを行なう。この方法では、計数操作が簡単である。

【００８７】これに対して、所定数以上の無効となったデータを含んでいるパリティグループの数をカウントして、そのカウント値が所定数を越えたときに、詰め替えを起動する方法でもよい。この方法では、詰め替えを行なう回数を制限できる。

【００８８】さらに別の方法として、空き領域の容量が所定値以下になったときに行なう方法でもよい。

【００８９】さらに、より望ましくは、ＣＰＵ１を介してユーザにより確保されたデータ書き込み容量のうち、空き領域の容量の占める割り合いが、所定値以下になったときに行なう方法でもよい。

【００９０】なお、書き替え領域に格納する段階で書き込み時と同様にアドレステーブル７０および動的アドレス変換テーブル９０を更新する。このようにして行内の全てのデータの無効フラグ９１がオンになると、その行を空き領域（書き替え領域）とする。

【００９１】以上の詰め変えは、以上の二つの方法のいずれを採るにしても論理グループ１０に発行される読み出し／書込み要求が比較的少ない時、特に書き込みが少ない時に行うとシステムの読み出し／書込み要求処理効率の低下が少なく効果的である。

【００９２】（１０）データの読み出し

ＭＰ２０が読み出し要求のコマンドを認識すると、ＣＰＵ１から指定されたＣＰＵ指定アドレス７１（ドライブ番号７４とＣＣＨＨＲ７５）からＭＰ２０はアドレステーブル７０を参照し、そのアドレスに対してＤＭポインタが登録されているか否かによりそのアドレスに対し動的アドレス変換を行なっているかを判断し、行なっていない場合は、引続きアドレステーブル７０により、当該

28

データのＳＣＳＩドライブアドレス７２へアドレス変換を行なう。同時にキャッシュメモリ７内に当該データが存在するかどうかアドレステーブル７０のキャッシュフラグ８２を調べ、判定する。一方、アドレステーブル７０のＤＭポインタ７６に登録されており、動的アドレス変換を行なっている場合は、動的アドレス変換テーブル９０により、アドレス変換を行なう。但し、無効フラグがオフであるドライブ内アドレスを利用する。これによりＣＰＵ指定アドレスが複数個テーブル９０に登録されている場合でも、最近に書き込まれたデータを読み出せるようになる。

【００９３】キャッシュヒット時は、ＭＰ２０はアドレステーブル７０又は動的アドレス変変換テーブル９０によりＣＰＵ１から指定してきたＣＰＵ指定アドレス７１（ドライブ番号７４とＣＣＨＨＲ７５）をキャッシュメモリ７のアドレス（キャッシュアドレス８１）に変換しキャッシュメモリ７へ当該データを読み出しに行く。具体的にはＭＰ２０の指示の元でキャッシュアダプタ回路２４によりキャッシュメモリ７から当該データは読み出される。キャッシュアダプタ２４はキャッシュメモリ７に対するデータの読みだし、書き込みをＭＰ２０の指示で行なう回路で、キャッシュメモリ７の状態の監視、各読みだし、書き込み要求に対し排他制御を行なう回路である。キャッシュメモリ７内には圧縮回路２７によりデータ圧縮されているデータも格納されているため、キャッシュアダプタ２４により読み出されたデータは圧縮回路２７に送られる。データ圧縮されていないデータは圧縮回路２７をそのまま通過するが、データ圧縮されているデータは、圧縮回路２７によりＣＰＵ１から転送されてきた圧縮前の元のデータに伸長される。圧縮回路２７を通過したデータはデータ制御回路（ＤＣＣ）２２の制御によりチャネルインターフェース回路２１に転送される。チャネルインタフェース２１ではＣＰＵ１におけるチャネルインターフェースのプロトコルに変換し、チャネルインターフェースに対応する速度に速度調整する。このプロトコル変換および速度調整後は、チャネルパスディレクタ５において、チャネルパススイッチ１６が外部インターフェースパス４を選択しインタフェースアダプタ１５によりＣＰＵ１へデータ転送を行なう。

【００９４】一方、キャッシュミス時はＭＰ２０はドライブインタフェース２８に対し、ＳＣＳＩドライブアドレス７２に従って当該ドライブ１２への読み出し要求を発行するように指示する。ドライブインタフェース２８ではＳＣＳＩインターフェースの読み出し処理手順に従って、読み出しコマンドをドライブユニットパス９－１、または９－２を介して発行する。ドライブインタフェース２８から読み出しコマンドを発行された当該ドライブ１２においては、指示されたＳＣＳＩドライブ内のアドレス（ＳＣＳＩ内Ａｄｄｒ７８）へシーク、回転待ちのアクセス処理を行なう。当該ドライブ１２における

29

アクセス処理が完了した後、当該ドライブ１２は当該データを読み出しドライブユニットバス９を介してドライブインタフェース２８へ転送する。ドライブインタフェース２８では転送されてきた当該データをドライブ側のキャッシュアダプタ回路１４に転送し、キャッシュアダプタ１４ではキャッシュメモリ７にデータを格納する。この時、このキャッシュアダプタ１４はキャッシュメモリ７にデータを格納することをＭＰ２０に報告し、ＭＰ２０はこの報告を元にアドレステーブル７０または動的アドレス変換テーブル９０のユーザが読みだし要求を発行したＣＰＵ指定アドレス７１に対応したキャッシュフラグ８２をオンにし、キャッシュメモリ７内のアドレスを以下のように更新する。ＣＰＵ１から読みだし要求先のアドレスとして指定してきた、ドライブ番号とＣＣＨＨＲに対し、アドレステーブル７０内のＣＰＵ指定アドレス７１（ドライブ番号７４とＣＣＨＨＲ７５）を探し、そのＣＰＵ指定アドレス７１に対応して、当該データを格納したキャッシュメモリ７内のアドレスをキャッシュアドレス８１に書き込む。これと同時にキャッシュフラグ８２をオンとする。キャッシュメモリ７にデータを格納し、アドレステーブル７０または動的アドレス変換テーブル９０のキャッシュフラグ８２をオンにし、キャッシュメモリ７内のアドレスを更新した後は、キャッシュヒット時と同様な手順でＣＰＵ１へ当該データを転送する。

【００９５】（１１）変形

本実施例ではパリティデータを作成する単位は、論理グループ１０を構成する各ＳＣＳＩドライブ１２について、同一アドレス（シリンダアドレス、ヘッドアドレス、レコードアドレスが全て等しい）とした。しかし、データ格納効率を向上させるため、同一アドレスに限定せず、インデックスからの距離が等しければ（レコードアドレスが等しい）シリンダアドレス、ヘッドアドレスは異なっていてもパリティデータを作成できる。このようにしても本実施例の動作は可能となる。

【００９６】（１２）実施例１のまとめ

以上述べたように、本実施例では、動的にアドレス変換をする領域へのデータの書き込みにあっては、複数の書き込みデータからパリティデータを生成し、それらの書き込みデータと生成したパリティデータとによりパリティグループを生成し、それらを空き領域に書き込む。この書き込みは、各書き込みデータがすでに書き込み済みのデータを更新するデータである場合にも適用される。

【００９７】従来のレベル５のＲＡＩＤによる書き込みのときには、書き込み済みのデータと書き込み済みの対応するパリティデータをドライブから読み出し、これらと更新用のデータとから、新たなパリティデータを生成し、この生成されたパリティデータとその更新用のデータとを旧のパリティデータが保持されている領域及び旧の書き込みデータが保持されている領域に書き込む。これらの二つの領域からの旧データの読み出し及びにそれらの領域への新データの書き込みのために、ドライブの回転待等の待時間がオーバヘッドとして大きい。

30

【００９８】本実施例では、以上に述べた書き込み方法により、以上の従来例で必要であった、ドライブからの旧データの読み出し、そこへの新データの書き込みが必要でない。そのため、以上のようなオーバヘッドが生じない。

【００９９】さらに、本実施例では、ある一つのパリティグループを構成する複数のデータの一部が書き換えられたとき、そのパリティグループのいずれのデータもドライブから読み出さない。つまり、書き換えられたデータは無効とし、それ以外のデータはそのまま有効なデータとしてドライブに保持する。この結果、一部のデータの書き換え時の処理が早くなる。

【０１００】さらに、本実施例では、一つのパリティグループを構成する複数のデータの長さを上位装置から送られてくるデータの一定長に等しくした。この結果、いずれかのパリティグループ内のデータが部分的に無効にされたとしても、そのパリティデータ内の全てのデータが無効となるケースが増大する。このように、完全に無効となった領域は空き領域として詰め替えをすることなくそのまま使用できる。従って、部分的に無効な状態のままでいるパリティグループは、より長いデータからパリティデータを構成した場合より少なくなり、結果として、詰め替えを要する部分的に無効なパリティグループの数は少なくなる。

【０１０１】さらに、本実施例では、一つのパリティグループを構成する複数のデータの長さを上位装置から送られてくるデータの一定長に等しくした結果、上位装置から転送された複数の書き込みデータを、それらからパリティグループを構成するまでの期間一時的に保持するためのキャッシュ内領域の大きさは小さくてよい。

【０１０２】（実施例２）本実施例では実施例１の図１に示すＡＤＣ２に対し、データの属性により動的にアドレス変換を行なうかをＡＤＣ２内で独自に判断する機能を付加する。本実施例ではシーケンシャルデータは、ＡＤＣ２のＭＰ１　２０が独自に判断して、動的なアドレス変換を行なう領域に格納するように制御する。実施例１では、ＣＰＵ１から指定されるアドレス（ＣＰＵ指定アドレス）７１をＭＰ２０はアドレステーブル７０または動的アドレス変換テーブル９０により、実際にデータの読みだしまたは書き込みを行なうＳＣＳＩドライブ１２に対するＳＣＳＩドライブアドレス７２に変換する。書き込み時はこの様なアドレス変換後、ＣＰＵ１から送られてきたデータを一端キャッシュメモリ７に格納する。その後は、キャッシュ独自のリプレースアルゴリズムでそのデータが追い出されるまで、そのデータをキャッシュに保持する。

【０１０３】本実施例では書き込みデータは全て一端キ

31

ャッシュメモリ７に格納し、一定時間キャッシュメモリ７内に保持する。このキャッシュメモリ７内に保持する時間は、ユーザがＡＤＣ２に対し予め設定することができ、ＡＤＣ２内のＭＰ２０により制御される。キャッシュメモリ７内に保持されたデータは、後述するように後に発行された書き込み要求のＣＰＵ指定アドレス７１とでＣＰＵ指定アドレス７１の比較を行ない、それぞれがシーケンシャルデータに属するかどうかを判断し、シーケンシャルデータに属するデータは動的アドレス変換を行なう領域に格納するようにＭＰ２０が独自に制御を行なう。

【０１０４】そこで、以下にＭＰ１　２０におけるシーケンシャルデータかどうかを判断する方法について示す。

【０１０５】書き込み時において、ＣＰＵ１からキャッシュメモリ７内へのデータの格納後、ＭＰ１　２０はＣＰＵ１から送られたＣＰＵ指定アドレス７１に対し、キャッシュメモリ７内のアドレス変換用テーブルの参照を行う。図３に示すように本実施例ではアドレステーブル７０の各ＣＰＵ指定アドレス７１に対し、アクセスフラグ８４を設定する。このアクセスフラグ８４は書き込み時においてＣＰＵ１が指定したＣＰＵ指定アドレス７１に対し、キャッシュメモリ７内への書き込みデータの格納後、ＭＰ１　２０がオン（１）とする。ユーザが予め指定した一定時間経過後、キャッシュメモリ７からＳＣＳＩドライブ１２へ格納した時点でＭＰ１　２０によりアクセスフラグ８４はオフ（０）とされる。

【０１０６】書き込み時において、ＣＰＵ１から指定されたＣＰＵ指定アドレス７１について、ＭＰ１　２０はアドレステーブル７０においてアクセスフラグ８４がオン（１）になっているＣＰＵ指定アドレス７１を調べ、ＣＰＵ１から指定されたＣＰＵ指定アドレス７１とアドレステーブル７０においてアクセスフラグ８４がオン（１）になっている当該ＣＰＵ指定アドレス７１を比較する。具体的には、ドライブ番号７４とＣＣＨＨＲ７５のシリンダアドレス（ＣＣ）を比較する。もし、前に発行され、キャッシュメモリ７内に格納されている（アクセスフラグ８４がオン（１））データのＣＰＵ指定アドレス７１と、後に発行された書込み要求のＣＰＵ指定アドレス７１の比較において、ドライブ番号７４が一致し、しかも、ＣＣＨＨＲ７５のシリンダアドレスが一致した場合、これらの書き込み処理はシーケンシャル処理と判定する。同様に次に発行された書込み要求についてもＣＰＵ指定アドレス７１を比較する。この様にして、発行されてきた書込み要求のデータについてＭＰ１　２０は、ユーザか予め設定しておいた一定時間の間シーケンシャル処理かどうかの判定を行なう。もし、シーケンシャルなデータの場合は、ＭＰ１　２０は独自にそのシーケンシャルなデータのグループに対し、当該ＳＣＳＩドライブ１２の動的なアドレス変換を行なう領域に書き

32

込む制御を行なう。具体的には、実施例１と同様にアドレステーブル７０および動的アドレステーブル９０の登録を行ない、キャッシュメモリ７から当該ＳＣＳＩドライブ１２に対する書込み処理をする。

【０１０７】（実施例３）本実施例では図８に示すように論理グループ１０単位にサブＤＫＣ１１を設け、その内部に実施例１、２において示したキャッシュメモリ７内のアドレステーブル７０および動的アドレス変換テーブル９０を持たせたものを示す。本実施例におけるデータの処理手順の内、実施例１および２で示した処理手順と異なる部分のみを図９および１０を用いて説明する。本実施例では図１０に示すように実施例１、２で示したキャッシュメモリ７内のアドレステーブル７０および動的アドレス変換テーブル９０を各論理グループ１０単位のサブＤＫＣ１１内のデータアドレステーブル（ＤＡＴ）３０に分割する。ＤＡＴ３０は格納されているテーブルの形式や機能は実施例１、２と同様であるが、異なるのは書き込み又は読み出しデータを格納するメモリとは別のアドレス変換用テーブルを格納する専用メモリ内に保持される点である。ＡＤＣ２内のキャッシュメモリ７内には、図５に示すようなグループアドレステーブル（ＧＡＴ）２３が格納されており、ＭＰ２０はこのＧＡＴ２３によりＣＰＵ１から指示されたＣＰＵ指定アドレス７１を、そのＣＰＵ指定アドレス７１が指示する場所がＡＤＵ３内のどの論理グループ１０かを判定する。

【０１０８】ＣＰＵ１からの読み出し要求が転送されたときにはＣＰＵ１から指定されたＣＰＵ指定アドレス７１により、ＭＰ２０はＧＡＴ２３で論理グループ１０を確定し、ＭＰ２０はこの当該論理グループ１０に対し読み出し要求を発行するようにドライブインタフェース２８に指示する。ＭＰ２０から指示を受けたドライブインタフェース２８は当該論理グループ１０のサブＤＫＣ１１に対し読み出し要求を発行する。サブＤＫＣ１１ではマイクロプロセッサＭＰ２９がこの読み出し要求のコマンドを受け付け、ＤＡＴ３０を参照し、実施例１で示したＭＰ２０がアドレステーブル７０および動的アドレステーブル９０を用いて処理したのと同様に当該データが格納されている論理グループ１０内のＣＰＵ指定アドレス７１に対するＳＣＳＩドライブアドレス７２およびパリティドライブアドレス７３を確定する。このアドレスの確定後、実施例１でのＭＰ２０の処理と同様に、ＭＰ２９は当該ＳＣＳＩドライブ１２に対し、読み出し要求を発行する。ＭＰ２９から読み出し要求を発行されたＳＣＳＩドライブ１２ではシーク、回転待ちを行ない、当該データの読み出しが可能になり次第当該データをドライブアダプタ回路３４に転送し、ドライブアダプタ３４はサブキャッシュメモリ３２に格納する。サブキャッシュメモリ３２に当該データの格納が完了した後、ドライブアダプタ３４はＭＰ２９に格納報告を行ない、ＭＰ２９は実施例１でのＭＰ２０と同様にＤＡＴ３０内のＣＰ

33

U指定アドレス７１に対応した当該キャッシュフラグ８２をオン（1）とし、キャッシュアドレス８１にサブキャッシュ３２内の格納したアドレスを登録する。後に当該キャッシュフラグ８２がオンのデータに対し読み出しまたは書込み要求が発行された場合は、サブキャッシュ３２内で処理を行なう。ＭＰ３　２９によるＤＡＴ３０の更新が終了すると、ＭＰ３　２９はＡＤＣ２内のＤｒｉｖｅ　ＩＦ２８に対しデータ転送可能という応答を行ない、ＤｒｉｖｅＩＦ２８はこの応答を受け取ると、ＭＰ１　２０に対し報告する。ＭＰ１　２０はこの報告を受け取ると、キャッシュメモリ７への格納が可能なら、Ｄｒｉｖｅ　ＩＦ２８に対しサブＤＫＣ１１からデータを転送するように指示する。Ｄｒｉｖｅ　ＩＦ２８ではＭＰ１　２０からの指示を受けるとサブＤＫＣ１１のＭＰ３　２９に対し読み出し要求を発行する。この読み出し要求を受けたＭＰ３　２９はサブキャッシュアダプタ回路（ＳＣＡ）３１に対しサブキャッシュ３２から当該データを読みだすように指示し、ＳＣＡ３１は実際にデータを読み出してＤｒｉｖｅ　ＩＦ２８にデータを転送する。Ｄｒｉｖｅ　ＩＦ　２８がデータを受け取った後は、実施例１、２で示した処理を行なう。

【０１０９】一方書込み時は読み出し時と同様に当該論理グループ１０を確定し、当該論理グループ１０のＭＰ３　２９に対し書込み要求を発行する。書込み要求を受け付け、書込みデータをサブキャッシュ３２に格納した後は、動的にアドレス変換して格納するデータについては、実施例１、２に示すように動的アドレス変換してドライブ１２に書込み、動的アドレス変換しない場合はレベル５の処理を行なう。なお、本実施例ではサブＤＫＣ１１内のサブキャッシュ３２に格納されたデータにおいてＰＧ３６がパリティデータを作成する。この様にサブＤＫＣ１１においてドライブ１２でのアドレス管理とデータ処理を行なうことにより負荷分散が図れる。

【０１１０】なお、図８では各論理グループ１０においてサブＤＫＣ１１が付いているが、一つのサブＤＫＣ１１が複数の論理グループ１０を管理しても構わない。このように複数の論理グループ１０に対しサブＤＫＣ１１をまとめておくことによりＤＡＴ３０の一括管理が行える。

【０１１１】なお、本実施例１、２は、磁気ディスクについて述べてきたが、光ディスク、フロッピディスクなどの記憶装置においても成り立つ。

【０１１２】

【発明の効果】以上述べたように、本発明に従って、複数の書き込みデータとそれらから生成されたパリティデータを含むパリティグループをドライブ内の空き領域に書き込む方法をとれば、データの書き換えを従来より高速に行なえる。

【０１１３】さらに、本発明に従って、ある一つのパリティグループを構成する複数のデータの一部が書き換えられたとき、書き換えられたデータは無効とし、それ以外のデータはそのまま有効なデータとしてドライブに保持するようにすれば、データの書き換え時の処理が早くなる。

34

【０１１４】さらに、本発明に従って、一つのパリティグループを構成する複数のデータの長さを上位装置より送られて来るデータの長さにすれば、いずれかのパリティグループ内のデータが部分的に無効にされたとしても、部分的に無効な状態のままでいるパリティグループは、より長いデータからパリティデータを構成した場合より少なくなり、結果として、詰め替えを要する部分的に無効なパリティグループの数は少なくなる。そのため、詰め替え処理の対象となるデータは少なくなる。さらに、上位装置から転送された複数の書き込みデータを、それらからパリティグループを構成するまでの期間一時的に保持するためのキャッシュ内領域の大きさは小さくてよい。

【図面の簡単な説明】

【図１】第１の実施例の全体構成図。

【図２】第１の実施例のクラスタ内構成図。

【図３】アドレステーブルの説明図。

【図４】動的アドレス変換テーブルの説明図。

【図５】第３の実施例で用いるグループアドレステーブル（ＧＡＴ）の説明図。

【図６】本発明のデータ格納方法の説明図。

【図７】本発明の詰め替え動作説明図。

【図８】第３の実施例の全体構成図。

【図９】第１の実施例のクラスタ内構成図。

【図１０】図９のサブＤＫＣ内構成図。

【図１１】従来のディスクアレイの説明図。

【図１２】従来技術の書込み時間の説明図。

【図１３】本発明のドライブの内部アドレス説明図。

【符号の説明】

２…アレイディスクコントローラ（ＡＤＣ）、３…アレイディスクユニット（ＡＤＵ）、４…外部インターフェースパス、５…チャネルパスディレクタ、６…チャネルパス、７…キャッシュメモリ、８…ドライブパス、９…アレイディスクユニットパス、１２…ドライブ、１３…クラスタ、１４…ドライブ側キャッシュアダプタ（Ｃ　Ａｄｐ）、１５…インターフェースアダプタ、１６…チャネルパススイッチ、１７…制御信号線、１８…データ線、１９…パス、２０…マイクロプロセッサ１（ＭＰ１）、２１…チャネルインターフェース（ＣＨ　ＩＦ）回路、２２…データ制御回路（ＤＣＣ）、２３…グループアドレス変換回路（ＧＡＴ）、２４…チャネル側キャッシュアダプタ（Ｃ　Ａｄｐ）、２７…圧縮回路、２８…ドライブインターフェース回路（Ｄｒｉｖｅ　ＩＦ）、２９…マイクロプロセッサ３（ＭＰ３）、３０…データアドレステーブル、３１…サブキャッシュアダプタ、３４…ドライブアダプタ（Ｄｒｉｖｅ　Ａｄｐ）、

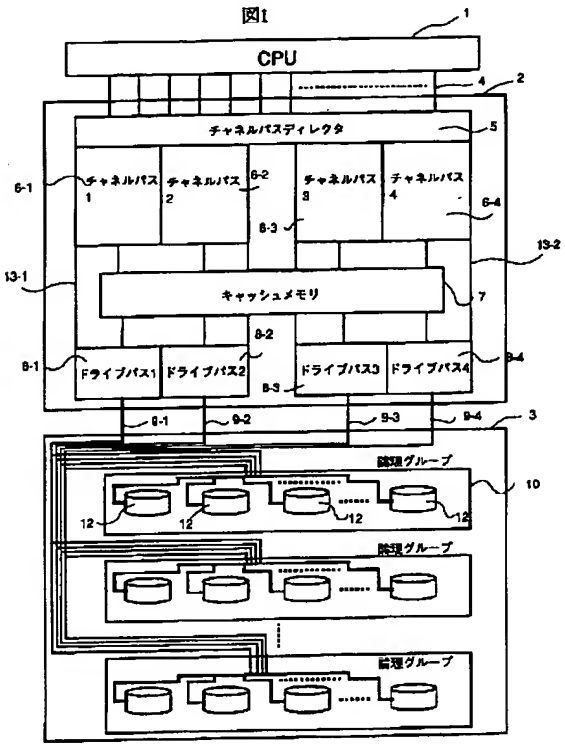
35

３５…ドライブパス、３６…パリティデータ生成回路、７０…アドレステーブル、７１…ＣＰＵ指定アドレス、７２…ＳＣＳＩドライブアドレス、７３…パリティドライブアドレス、７４…ドライブ番号、７５…ＣＣＨＨＲ、７６…ＤＭポインタ、７７…ＳＣＳＩドライブ番号、７８…ＳＣＳＩドライブ内アドレス、７９…パリティドライブ番号、８０…パリティドライブ内アドレス、８１…キャッシュアドレス、８２…キャッシュフラグ、８３…ドライブフラグ、８４…アクセスフラグ、８５…論理グループアドレス、９０…動的アドレス変換テーブル、１００…グループアドレステーブル（ＧＡＴ）。

36

【図１】

図1



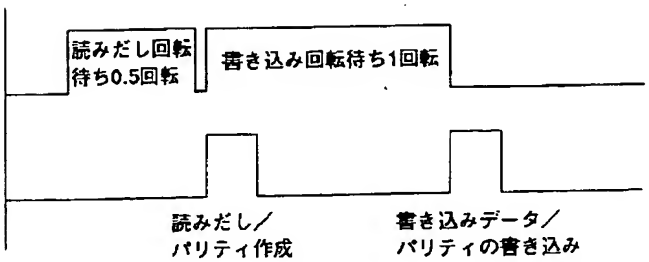
【図５】

図5

CPU指定アドレス				
ドライブ番号	CCHHR	論理グループアドレス	キャッシュアドレス	Cache Flag
Drive#1	ADR 1	LADR 1	——	——
	ADR 2	LADR 3	——	——
	ADR 3	LADR 8	——	——
	⋮	⋮	⋮	⋮
Drive#2	ADR 1	LADR 2	——	0
	ADR 2	LADR 1	CADR1,5	1
	ADR 3	LADR 5	CADR1,8	1
	ADR 4	LADR 4	CADR1,6	1
⋮	⋮	⋮	⋮	⋮

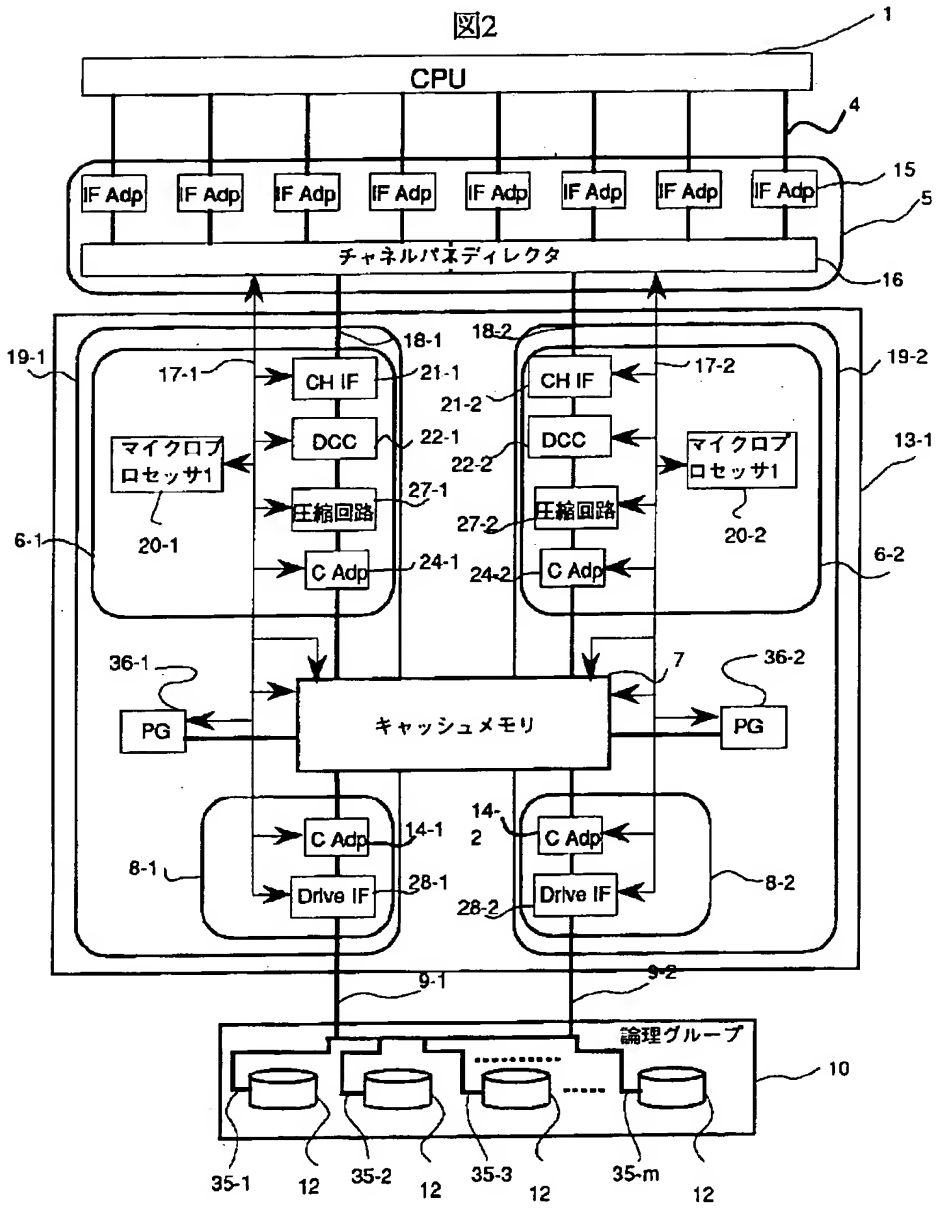
【図１２】

図12



【図2】

図2


1
CPU
4
IF Adp
15
5
チャネルパスディレクタ
16
18-1
18-2
19-1
19-2
17-1
CH IF
21-1
21-2
17-2
DCC
22-1
22-2
マイクロプロセッサ1
13-1
20-1
20-2
27-1
27-2
圧縮回路
6-1
6-2
24-1
24-2
C Adp
36-1
36-2
7
PG
キャッシュメモリ
14-1
14-2
8-1
8-2
28-1
28-2
Drive IF
9-1
9-2
論理グループ
10
35-1
35-2
35-3
35-m
12

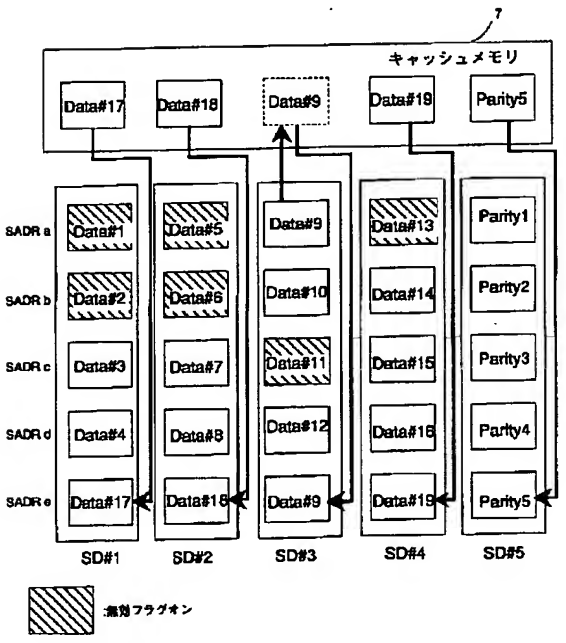
【図3】

図3

74	CPU指定アドレス			SCSI Driveアドレス		Parity Driveアドレス				
	ドライブ番号	CCHHR	DMポインタ	SCSI Drive No	SCSI内Addr	Parity Drive No	Parity内Addr	キャッシュアドレス	Cache Flag	アクセスフラグ
		ADR 1	DM a-2	——	——	——	——	——	——	1
		ADR 2	DM c-3	——	——	——	——	——	——	0
	Drive#1	ADR 3	DM d-2	——	——	——	——	——	——	1
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		ADR 1	——	SD#8	SADR 1	SD#12	SADR 1	——	0	1
		ADR 2	——	SD#9	SADR 1	SD#12	SADR 1	CADR1,5	1	0
	Drive#2	ADR 3	——	SD#10	SADR 1	SD#12	SADR 1	CADR1,8	1	1
		ADR 4	——	SD#11	SADR 1	SD#12	SADR 1	CADR1,6	1	1
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

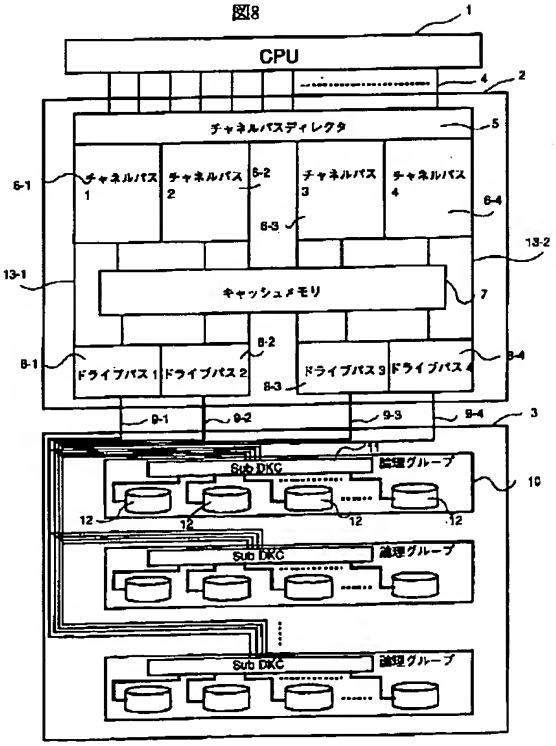
【図7】

図7



【図8】

図8



【図4】

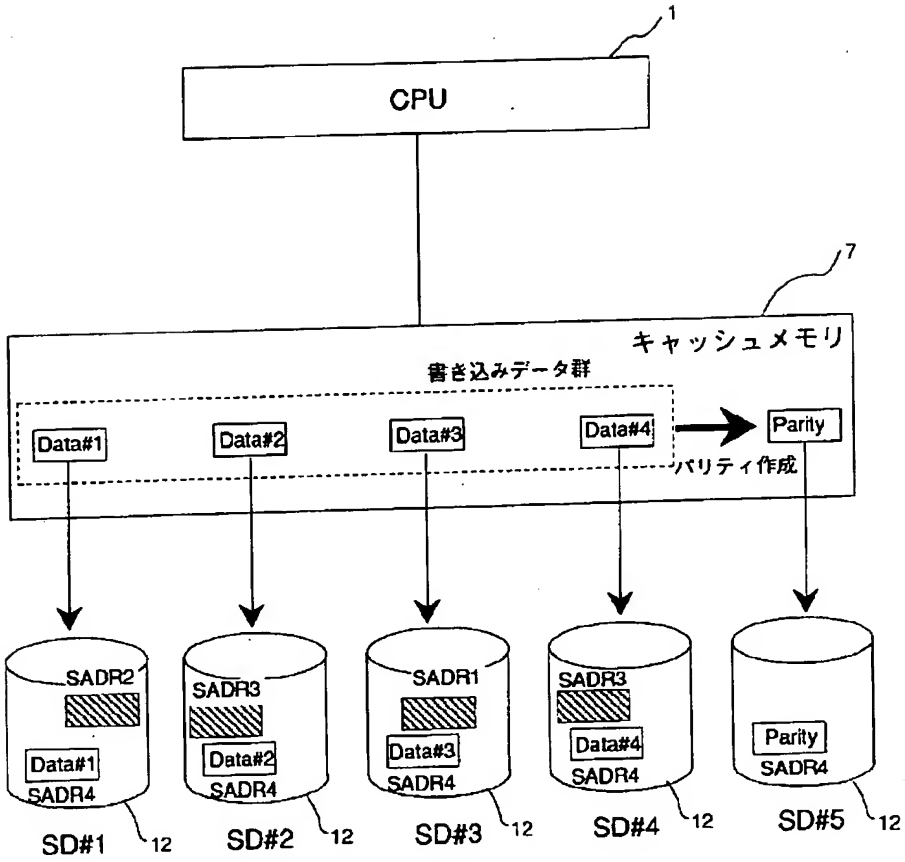
図4

DMポインタ (76)	SCSI Driveアドレス		無効Flag (91)	Rarity Driveアドレス		キャッシュアドレス (81)	Cache Flag (82)	Drive Flag (83)
	SCSI Drive No (77)	SCSI 内 Addr (72, 78)		Parity Drive No (79)	Parity内 Addr (73, 80)			
DM a-1	SD#1	SADR 1	0	SD#5	SADR 1	CADR2.5	1	1
DM a-2		SADR 2	1	SD#5	SADR 2	CADR2.8	1	0
DM a-3		SADR 3	1	SD#5	SADR 3	——	0	0
DM a-4		SADR 4	1	SD#5	SADR 4	——	0	0
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮
DM b-1	SD#2	SADR 1	0	SD#5	SADR 1	——	0	0
DM b-2		SADR 2	1	SD#5	SADR 2	——	0	0
DM b-3		SADR 3	1	SD#5	SADR 3	CADR2.3	1	0
DM b-4		SADR 4	1	SD#5	SADR 4	——	0	0
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮
DM c-1	SD#3	SADR 1	1	SD#5	SADR 1	CADR2.9	0	0
DM c-2		SADR 2	1	SD#5	SADR 2	——	0	0
DM c-3		SADR 3	0	SD#5	SADR 3	——	0	1
DM c-4		SADR 4	1	SD#5	SADR 4	——	0	0
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮
DM d-1	SD#4	SADR 1	1	SD#5	SADR 1	——	0	0
DM d-2		SADR 2	0	SD#5	SADR 2	CADR2.2	1	1
DM d-3		SADR 3	1	SD#5	SADR 3	CADR2.7	1	0
DM d-4		SADR 4	1	SD#5	SADR 4	——	0	1
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮
DM e-1	SD#5	SADR 1	0	SD#5	SADR 1	——	0	0
DM e-2		SADR 2	0	SD#5	SADR 2	——	0	1
DM e-3		SADR 3	0	SD#5	SADR 3	——	0	0
DM e-4		SADR 4	0	SD#5	SADR 4	——	0	0
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮

(90)

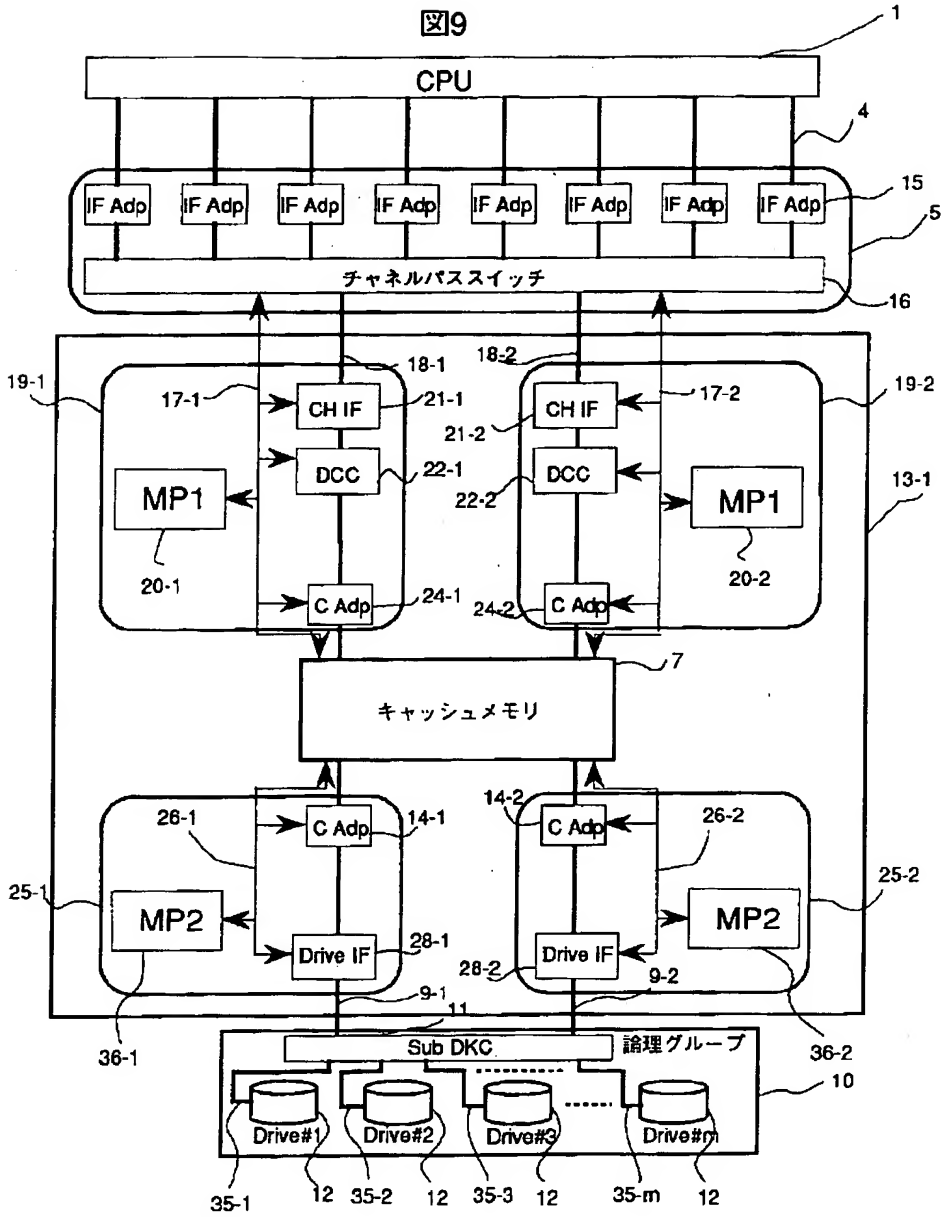
【図6】

図6


1
CPU
7
キャッシュメモリ
書き込みデータ群
Data#1
Data#2
Data#3
Data#4
Parity
パリティ作成
SADR2
Data#1
SADR4
SADR3
Data#2
SADR4
SADR1
Data#3
SADR4
SADR3
Data#4
SADR4
Parity
SADR4
SD#1
12
SD#2
12
SD#3
12
SD#4
12
SD#5
12
:CPU指定アドレスに対応する動的アドレス変換前SCSIドライブアドレス
:CPU指定アドレスに対応する動的アドレス変換後SCSIドライブアドレス

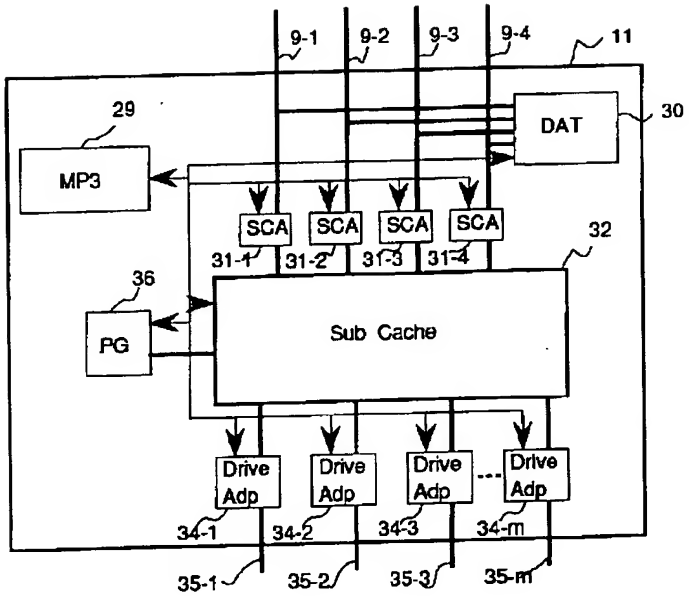
【図9】

図9


1
CPU
4
IF Adp
IF Adp
IF Adp
IF Adp
IF Adp
IF Adp
IF Adp
IF Adp
15
5
チャネルパススイッチ
16
18-1
18-2
19-1
17-1
CH IF
21-1
CH IF
17-2
19-2
21-2
DCC
22-1
DCC
22-2
MP1
MP1
13-1
20-1
20-2
C Adp
24-1
C Adp
24-2
7
キャッシュメモリ
26-1
C Adp
14-1
14-2
C Adp
26-2
25-1
MP2
MP2
25-2
Drive IF
28-1
Drive IF
28-2
9-1
9-2
11
36-1
Sub DKC
論理グループ
36-2
10
Drive#1
Drive#2
Drive#3
Drive#m
35-1
12
35-2
12
35-3
12
35-m
12

【図１０】

図10


9-1
9-2
9-3
9-4
11
29
MP3
DAT
30
SCA
SCA
SCA
SCA
31-1
31-2
31-3
31-4
32
36
PG
Sub Cache
Drive Adp
Drive Adp
Drive Adp
Drive Adp
34-1
34-2
34-3
34-m
35-1
35-2
35-3
35-m

【図１３】

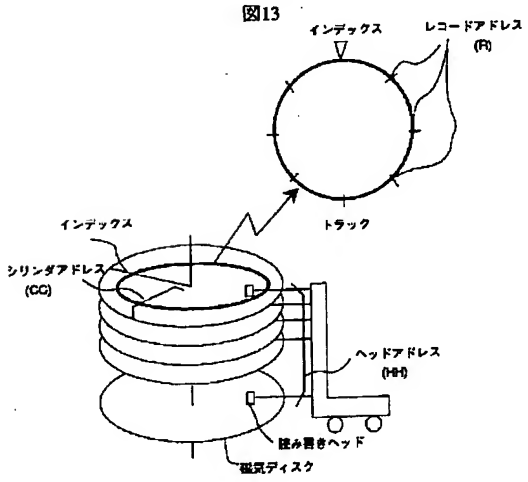
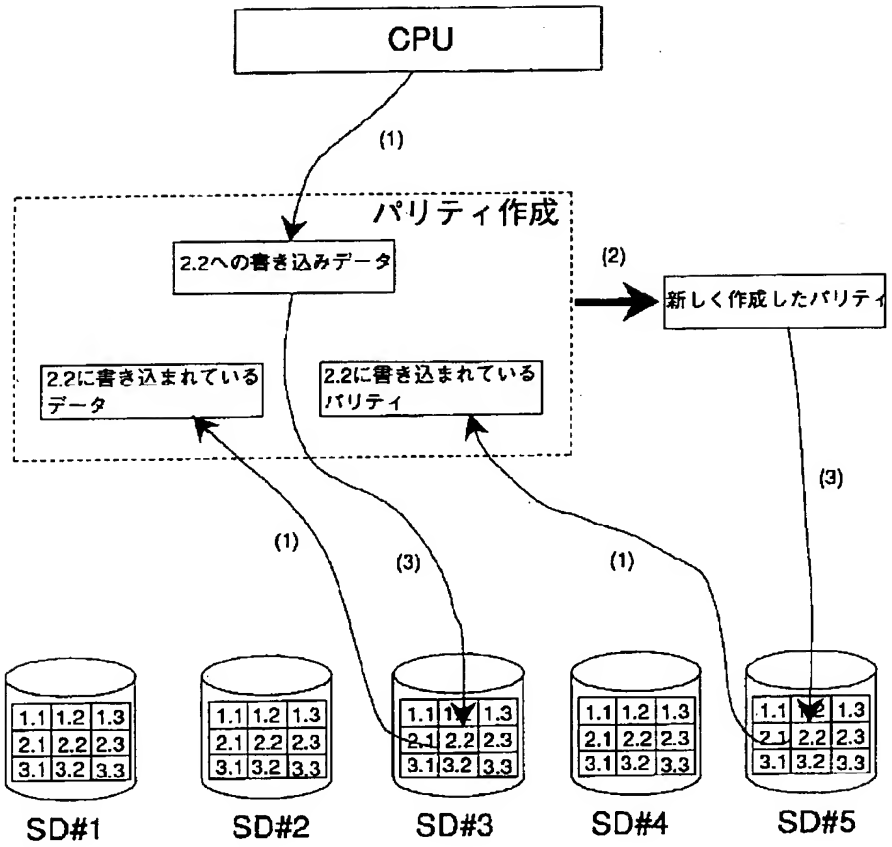


図13
インデックス
レコードアドレス
(R)
トラック
インデックス
シリンダアドレス
(CC)
ヘッドアドレス
(HH)
読み書きヘッド
磁気ディスク

【図11】

図11


CPU
(1)
パリティ作成
2.2への書き込みデータ
(2)
新しく作成したパリティ
2.2に書き込まれているデータ
2.2に書き込まれているパリティ
(3)
(1)
(3)
(1)
1.1 1.2 1.3
2.1 2.2 2.3
3.1 3.2 3.3
SD#1
SD#2
SD#3
SD#4
SD#5